THE ASTROPHYSICAL JOURNAL, 535:190–210, 2000 May 20

SINGULAR ISOTHERMAL DISKS. I. LINEAR STABILITY ANALYSIS

FRANK H. SHU AND GREGORY LAUGHLIN
Department of Astronomy, University of California Berkeley, CA 94720; shu@vshu.berkeley.edu; gpl@astro.berkeley.edu
SUSANA LIZANO
Institute de Astronomía, UNAM Apdo 70-264 4510 México, D.F., Mexico; lizano@astrosmo.unam.mx
AND
DANIELE GALLI
Osservatorio Astrofisico di Arcetri Largo Enrico Fermi, 5 I-50125 Firenze, Italy; galli@arcetri.astro.it
Received 1999 September 14; accepted 2000 January 3

ABSTRACT

As part of a larger effort to understand how binary and single stars form from the collapse of magnetized molecular cloud cores, we perform a global stability analysis of isopedically magnetized, singular isothermal disks (SIDs). The work described here has precedents in earlier studies of disturbances in power-law disks by Zang in 1976, Toomre in 1977, Lynden-Bell & Lemos in 1993, Syer & Tremaine in 1996, and Goodman & Evans in 1999. We find the analytic criteria for the bifurcation of axisymmetric disks into nonaxisymmetric forms with azimuthal periodicities $m = 1$ (eccentric displacements), 2 (oval distortions), 3 (triangular distortions), etc. These bifurcations, which occur at zero frequency, are the compressible and differentially rotating analogs of how the classical sequence of incompressible and uniformly rotating Maclaurin spheroids bifurcate (secularly, under dissipative forces) to become Dedekind ellipsoids with figure axes that remain fixed in space. Like Syer & Tremaine and Lynden-Bell & Lemos, we also find that zero-frequency logarithmic spirals are possible scale-free disturbances, but our interpretation of the existence of such steadily propagating wavetrains is different. We give a dynamical instability interpretation based on the onset of swing amplification by overreflection at the corotation circle of prograde spiral density waves the pattern speeds of which have nonzero and positive values. Our analysis yields identical instability criteria as the global normal-modes treatment of Goodman & Evans, and we tentatively also identify dynamical barred-spiral instabilities as the "breathing mode" limit of two-armed ordinary-spiral instabilities. We prove a general "reciprocity theorem," which states that the overreflection factors are *identical* for spiral density waves launched from cavities interior or exterior to Q-barriers that straddle the corotation circle. This globally valid result supports a unifying interpretation, advocated for many years by C. C. Lin and his colleagues (see, e.g., work by Bertin & Lin): the coexistence of spiral structure in galaxies arising from the instability of internal normal modes in the combined star/gas disk or from driving by external tidal influences associated with the chance passages of companion bodies.

Subject headings: circumstellar matter — galaxies: spiral — hydrodynamics — instabilities — solar system: formation

1. INTRODUCTION

1.1. *Previous Work*

Since Mestel's (1963) pioneering study of singular isothermal disks (SIDs) and their relatives, a growing astrophysical interest has arisen, in both the star-formation and galactic-structure communities, in the dynamics of self-gravitating configurations with radial profiles of surface density Σ or volume density ρ that are power laws in cylindrical radius ϖ or spherical radius r, i.e., in systems that are scale-free. The potential applications range from the structure of barred and spiral galaxies (see, e.g., Bertin & Lin 1996) to the light cusps seen in the nuclei of galaxies (see, e.g., Crane et al. 1993) to the formation and collapse of cloud cores in the birth of stars and planetary systems (see, e.g., Shu et al. 1999).

Self-similar equilibria and collapses have been studied in two and three spatial dimensions, usually but not always with the simplifying assumption of axial symmetry. Thus, Shu (1977) considered the self-similar gravitational collapse of the gaseous, singular isothermal sphere (SIS with $\rho \propto r^{-2}$). Toomre (1982) and Hayashi, Narita, & Miyama (1982) constructed, respectively, self-consistent, axisymmetric, three-dimensional (i.e., unflattened) equilibria of "isothermal" stellar and gaseous systems with flat rotation curves (with $\rho \propto r^{-2}$ times a function of colatitude θ). Baureis, Ebert, & Schmitz (1989) generalized the nonrotating, gaseous SIS to include the effect of scale-free frozen-in magnetic fields. Following a long line of research on molecular cloud core formation by ambipolar diffusion (e.g., Nakano 1979; Lizano & Shu 1989; Basu & Mouschovias 1994), Li & Shu (1996) argued that the process naturally tends to produce pivotal states just before dynamical collapse in which the central regions acquire mass-to-flux ratios that are constant and have the density distributions of singular isothermal toroids [SITs with $\rho \propto r^{-2}R(\theta)$ where $R(\theta)$ vanishes at the poles], thereby unknowingly reproducing and extending the results of Baureis et al.

A breakthrough was achieved recently in this field by Syer & Tremaine (1996), who obtained semianalytic descriptions of nonaxisymmetric equilibria for completely flattened (razor-thin) power-law disks. Unfortunately, their method is unsystematic with regard to finding nonlinear

equilibria of nonaxisymmetric form. By utilizing Kalnajs's (1971) formulation of density-potential pairs decomposed as logarithmic spirals to describe all disturbances, it also obscures the physical differences between aligned (i.e., nonspiral) and unaligned (i.e., spiral) perturbations. Part of the purpose of the present series of papers is to supplement the analysis of Syer & Tremaine by introducing alternative calculational techniques in the special case of a singular isothermal disk where the equilibrium surface density $\Sigma \propto \varpi^{-1}$.

Self-gravitating disturbances of a barlike or spiral-like form, driven either internally or externally to the system, have long been regarded as important for the inward transport of mass and the outward transport of angular momentum in disk galaxies and planetary rings (e.g., Toomre 1969; Shu 1970; Lynden-Bell & Kalnajs 1972; Huntley, Sanders, & Roberts 1978; Goldreich & Tremaine 1978, 1979). They have been hypothesized (e.g., Laughlin, Korchagin, & Adams 1998) to be equally important in the early evolution of the protoplanetary disks that have been imaged around young stellar objects (e.g., Skrutskie et al. 1993; Kawabe et al. 1993; Lay et al. 1994). Interstellar magnetic fields dragged into the disk from the collapse of the star-forming cloud core would generally complicate the dynamics of such protoplanetary disks, yielding among other possibilities the magnetorotational instability (e.g., Balbus & Hawley 1991) or even magneto-centrifugally driven disk winds (e.g., Königl & Pudritz 2000). However, in its effect on the modification of self-gravitating disturbances, the results are easy to calculate in the case when the disk is *isopedically* magnetized, i.e., when the local mass-to-flux ratio λ has a spatially uniform value and when field-freezing makes λ also temporally constant. If we adopt the isopedic assumption, together with the model that the field outside the disk exists effectively in a vacuum (no electrical currents), the theorems of Shu & Li (1997) allow us to incorporate the effects of frozen-in magnetic fields inside the disk by the simple expedient of rescaling from an unmagnetized disk.

The study of nonaxisymmetric perturbations in the linearized regime can also benefit from the adoption of basic states that are power laws. For example, Zang (1976; see the summary in Toomre 1977) studied the stability properties of an infinitesimally thin axisymmetric disk of stars with a flat rotation curve and uniform velocity dispersion (a stellar-dynamical SID with $\Sigma \propto \varpi^{-1}$). On dimensional grounds, Zang reasoned that no true normal modes were possible unless modifications (such as central cut-outs) were introduced to break the self-similarity of the SID. Evans & Read (1998) adopted Zang's method to examine discrete, growing, normal modes in stellar-dynamical disks of general power-law forms with cut-out centers. In contrast, Lynden-Bell & Lemos (1991, unpublished, but posted on the preprint server in 1993) argued that while compressible disks the equilibria of which are simple power laws probably have no unstable discrete modes, they do possess a continuum that becomes unstable simultaneously under changing conditions of the equilibrium disk. Goodman & Evans (1999) claimed that discrete normal modes could be defined, even for unmodified gaseous SIDs, provided that one specifies the phase of a postulated reflection of spiral waves from the origin. Goodman & Evans derived distinct criteria for the onset of instability for individual modes, thereby arriving at conclusions seemingly contradictory to those reached by Lynden-Bell & Lemos.

1.2. *Accomplishments of This Work*

In our analysis, aligned disturbances represent scale-free and zero-frequency bifurcations to alternative states of equilibria. These alternative states are in some sense preferred energetically since they are more compact than their axisymmetric counterparts; however, Kelvin's circulation theorem prevents their being reached by inviscid evolution from other states unless the two states have the same ratio $\mathscr{C}$ of circulation to enclosed mass. To the numerical accuracy of our calculations, $\mathscr{C}$ is very nearly a constant along the barlike ($m = 2$) bifurcation sequence. Thus, in principle, it is possible to evolve along the sequence, with little or no redistribution of circulation, as long as one can dissipate the energy difference between the states, which is also relatively small before the sequence terminates in shock waves (Paper II, forthcoming, will discuss this). For lack of a better terminology, we shall therefore refer to the aligned bifurcations as secular instabilities. The looseness of this terminology is not critical since the axisymmetric and nonaxisymmetric singular equilibria in all the bifurcation sequences are unstable to gravitational collapse. Thus, left on their own, they could never evolve from one unstable equilibrium to another along a line of relatively weak instability but would take a much faster dynamical route to yet more compact states. To reach the nonaxisymmetric equilibria realistically would probably require dissipative evolution from nonsingular states of quasi-magnetostatic equilibria that have not yet developed density cusps at their centers. Unlike the corresponding Maclaurin-Jacobi or Maclaurin-Dedekind bifurcations (see Chandrasekhar 1969), the most important dissipative agent in the star formation context is neither viscosity nor gravitational radiation, but probably ambipolar diffusion or turbulent decay (see Shu et al. 1999).

In contrast to the aligned disturbances, the unaligned or spiral disturbances, even at zero frequency, generally possess a different power-law scaling than the axisymmetric background (a property recognized explicitly by Syer & Tremaine 1996). They are the limiting example of a family of prograde, swing-amplifying (see Goldreich & Lynden-Bell 1965; Julian & Toomre 1966; Toomre 1969, 1977; Mark 1976; and Bertin et al. 1989), spiral density waves that are possible in almost all self-gravitating thin disks. For such propagating waves to establish a true normal mode (either growing or not growing secularly in time), they must be able to reflect off the center of the configuration (or some other inner "barrier"). At zero frequency, the spiral disturbances assume a logarithmic (or equiangular) pattern. Logarithmic spiral disturbances take infinitely many wavelengths to reach the origin (a problem already recognized by Zang 1976; see also Toomre 1977); thus, such disturbances do not naturally form coherent superpositions of trailing and leading patterns. To find spiral normal modes, Zang needed to cut a "hole" in his stellar SID, which not only removed a single logarithmic spiral as an eigenfunction but also introduced a length scale that combined with the velocity scale of a flat rotation curve to define a nonzero wave frequency and growth rate for a true normal mode (a standing pattern of unequal leading and trailing spirals).

Goodman & Evans (1999) pointed out that an ingoing wavetrain, even one containing an infinite number of wavelengths, has a constant, nonzero, group velocity in an unmodified SID and reaches the origin in *finite* time; therefore such a wavetrain cannot simply disappear but must

reflect from a trailing-spiral waveform to a leading-spiral waveform. Unfortunately, this reflection occurs with an ill-determined change of phase in power-law disks because the perturbation surface density contains too strong a singularity as $\varpi \to 0$ (signaling a breakdown of the basic assumptions of a linearized stability analysis) for any combination of leading and trailing spirals to cancel out in superposition and give a regular solution at the origin. Adopting instead an arbitrary prescription for the phase relationship between leading and trailing spirals, Goodman & Evans were able to show that an unmodified SID could support a discrete spectrum of normal modes and to compute the ratios of the resulting growth rates to pattern speeds given the properties of the axisymmetric equilibria.

In this paper, we shall resolve the discrepancies between the disparate viewpoints, not by studying the full problem of growing-spiral normal-modes that includes reflections at inner or outer boundaries, but by studying only the portion of it that involves wave propagation and amplification. In particular, we investigate the swing process that occurs when leading trailing waves impinge on a corotation circle, either from the inside or the outside, and transmit and overreflect into trailing spiral waves on either side of corotation (see, e.g., Mark 1976). For a given equilibrium model and azimuthal quantum number $m \neq 0$, we find that wavetrains of arbitrary (but positive) pattern speed, ω/m, have *identical* overreflection factors in SIDs, thus justifying Lynden-Bell & Lemos's claim that growth can set in for a continuum of disturbances simultaneously. However, the presence of a universal wave amplifier does not by itself allow the presence of normal modes that grow in time. For that to happen, coherent feedback must occur via reflecting boundary conditions to establish a resonant cavity (see Fig. 12.4 of Shu 1992). To obtain such reflections naturally, and not by fiat, it is necessary to remove the singularities of the pure power-law disk, for example, by inner cut-outs that mimic in a spiral galaxy the effects of a transition from disk to bulge. The "quantum condition" associated with the cavity condition then selects a discrete number of pattern speeds out of the continuum of possibilities for actual coherent amplification (see Bertin et al. 1989). Lynden-Bell & Lemos err, nevertheless, when they conclude that zero-frequency disturbances do not signal the onset of instability. By taking $\omega = 0$ before considering perturbations, they move the corotation circle to infinity, where it is unable to take part in the wave amplification process. We obtain the opposite answer when we perform the two limiting operations in a different order. Indeed, our criterion for the onset of overreflection makes intimate use of the properties of zero-frequency disturbances, and it is in precise agreement with Goodman & Evan's criterion for instability from their normal-modes treatment.

In summary, the resolution of the differences of opinion among Zang, Lynden-Bell & Lemos, and Goodman & Evans has two sources. Consideration only of the swing amplification process across corotation allows simultaneous wave growth in power-law disks for a continuum of disturbances. Consideration also of realistic boundary conditions (not done in this paper) selects from this continuum of possibilities a discrete spectrum of instabilities involving normal modes that grow coherently in time (and not just in space). Indeed, power-law disks provide a powerful vehicle for cleanly separating the study of the first element (propagation and amplification) from the second (feedback) since we can treat the global process of swing amplification in a whole disk (automatically excluding reflections at inner and outer boundaries) rather than in a truncated description involving, say, the approximation of a "swinging sheet" (see Goldreich & Lynden-Bell 1965 or Julian & Toomre 1966).

The rest of this paper is organized as follows. In § 2, we review the equations underlying nonaxisymmetric disturbances in SIDs. In § 3, we derive the rotation rates at which axisymmetric SIDs can bifurcate to nonaxisymmetric forms with aligned figure axes that are fixed in space for arbitrary values of the azimuthal quantum number m. In § 4, we analyze the properties of spiral density waves with zero pattern speed. The resulting exact dispersion relation for logarithmic spirals allows us to find the rotation rate of the underlying disk needed to support disturbances of a particular m-fold symmetry and winding angle. In § 5, we consider spiral disturbances with positive pattern speed, where a single logarithmic spiral no longer contains the correct spatial variation, and we must solve the governing linearized equation numerically to obtain the overreflection by swing amplification that occurs when steady wavetrains interact across a corotation circle. These results motivate our derivation in § 6 of a "reciprocity theorem." In § 7, we give a pictorial depiction of wave propagation, tunneling and transmission across Q-barriers, and overreflection by a time-dependent study of wave packets rather than a time-independent study of steady wavetrains. Finally, in § 8, we give a physical summary of the implications of our work.

2. MAGNETIZED SIDS

We use a cylindrical coordinate system (ϖ, φ, z) to describe an electrically conducting, infinitesimally thin, disk located in the plane $z = 0$ and threaded by a magnetic field that is assumed to exist for $|z| > 0$ as a vacuum field that fans out above and below but does not return to the disk. Because our model disks are supposed to represent, as our main interest, either idealized versions of molecular cloud cores or protostellar disks that form during an early stage of collapse from such cloud cores before much flux loss has occurred, we approximate them to be isopedic, with a dimensionless ratio λ of the mass per unit area $\Sigma(\varpi, \varphi, t)$ to flux per unit area $B_z(\varpi, \varphi, t)$ (perpendicular component of the magnetic field evaluated at $z = 0$), measured in units of $(2\pi G^{1/2})^{-1}$, that is spatially constant:

$$\lambda = 2\pi G^{1/2} \frac{\Sigma}{B_z} = \text{const} . \tag{1}$$

For gravitationally bound and physically isolated disks subject to free boundary conditions, $|\lambda| \geq 1$.

In isopedic SIDs, a frozen-in magnetic field adds no extra complication to a hydrodynamical treatment. The equations of motion for the isopedic, isothermal SID are identical with those of a nonmagnetized disk except for the transformations (Shu & Li 1997)

$$a^2 \to \Theta a^2, \quad G \to \epsilon G ,$$

where ϵ is the dilution factor for self-gravity caused by magnetic tension,

$$\epsilon = 1 - \frac{1}{\lambda^2} , \tag{2}$$

and Θ is the enhancement factor for gas pressure caused by magnetic pressure,

$$\Theta = \frac{\lambda^2 + 1 + 2\eta^2}{\lambda^2 + \eta^2} . \tag{3}$$

In the above, a is the isothermal sound speed, and η is defined to be the ratio of the horizontal gravitational field $g_{\|}$ (in absolute value) to the vertical gravitational field $2\pi G\Sigma$ (in absolute value) just above or below the disk,

$$\eta \equiv \frac{|g_{\|}|}{2\pi G\Sigma} . \tag{4}$$

With $\lambda \geq 1$, we have $0 \leq \epsilon \leq 1$ and $1 \leq \Theta \leq 2$. For an unmagnetized SID, we have $\epsilon = 1$ and $\Theta = 1$.

2.1. *Basic Equations*

We denote u, j, and $\Pi_g = a^2\Sigma$ as, respectively, the ϖ-component of the fluid velocity, the z-component of the specific angular momentum, and the vertically integrated gas pressure. The continuity and momentum equations for an inviscid, perfectly conducting, fluid can now be written as

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{\varpi}\frac{\partial}{\partial \varpi}(\varpi \Sigma u) + \frac{1}{\varpi^2}\frac{\partial}{\partial \varphi}(\Sigma j) = 0 ; \tag{5}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial \varpi} + \frac{j}{\varpi^2}\frac{\partial u}{\partial \varphi} - \frac{j^2}{\varpi^3} = -\frac{1}{\Sigma}\frac{\partial}{\partial \varpi}(\Theta \Pi_g) - \epsilon\frac{\partial \mathscr{V}}{\partial \varpi} ; \tag{6}$$

$$\frac{\partial j}{\partial t} + u\frac{\partial j}{\partial \varpi} + \frac{j}{\varpi^2}\frac{\partial j}{\partial \varphi} = -\frac{1}{\Sigma}\frac{\partial}{\partial \varphi}(\Theta \Pi_g) - \epsilon\frac{\partial \mathscr{V}}{\partial \varphi} ; \tag{7}$$

where the gravitational potential, $\mathscr{V}$ is given by Poisson's integral,

$$\mathscr{V}(\varpi, \varphi, t) = -G \oint d\psi \times \int_0^\infty \frac{\Sigma(r, \psi, t) r\, dr}{[r^2 + \varpi^2 - 2r\varpi \cos(\psi - \varphi)]^{1/2}} . \tag{8}$$

2.2. *Axisymmetric Equilibrium State*

We adopt an axisymmetric SID as a reference state, which has the following equilibrium properties:

$$\Sigma_0 = \frac{\Theta a^2}{2\pi\epsilon G\varpi}(1 + D^2) ,$$

$$\Omega \equiv \frac{j_0}{\varpi^2} = \frac{\Theta^{1/2} a}{\varpi} D ,$$

$$\kappa \equiv \left(\frac{2\Omega}{\varpi}\frac{dj_0}{d\varpi}\right)^{1/2} = \sqrt{2}\,\frac{\Theta^{1/2} a}{\varpi} D . \tag{9}$$

In the above, D is a dimensionless constant characterizing the level of rotation; Ω and κ are, respectively, the circular and epicyclic frequencies; and we have assumed that Θ is a constant. This is a valid assumption in the axisymmetric equilibrium state because $|g_{\|}| = |d\mathscr{V}_0/d\varpi| = \Theta a^2(1 + D^2)/\epsilon\varpi = 2\pi G\Sigma_0$, i.e., $\eta = 1$. With this result,

$$\Theta = \frac{\lambda^2 + 3}{\lambda^2 + 1} \tag{10}$$

is constant when λ is constant. In what follows, we assume for simplicity that Θ remains a constant even in the presence of perturbations in $g_{\|}$ and Σ. For practical applications, little is gained by taking a more sophisticated treatment of the magnetic enhancement of the gaseous contribution.

2.3. *Perturbations Periodic in Angle and Time*

On dimensional grounds, a SID does not support true normal modes with a time dependence proportional to $e^{i\omega t}$, except for the special case of zero-frequency modes, $\omega = 0$ (see also Lynden-Bell & Lemos 1993). The zero-frequency modes are generally interpreted either as bifurcation points (the start of a nonaxisymmetric sequence of equilibria) or as points of marginal stability (crossover points in normal-mode behavior from solutions that decay as a function of time to those that grow in time, probably as power laws since oscillatory and exponential behaviors are not allowed). The situation is more ambiguous for singular configurations because the homogeneous boundary conditions usually attached to a normal-modes approach (to assure that waves are not fed in from the boundaries) are more difficult to apply because of the singularities there.

To make the situation clear, we develop the formalism as if disturbances could have dependences $\propto e^{i\omega t}$ and provide the physical interpretations later (see below), both for the limit $\omega \to 0$ and for the case $\omega \neq 0$. We shall find that spiral density waves, emitted from one boundary of the system (origin or infinity) and propagating indefinitely long toward the other, can be amplified en route (by overreflection) if a corotation circle exists across which waves of opposite signs of angular momentum (and energy) densities can interact (see Shu 1992, chap. 12). The criterion for the existence of a corotation circle in an unbounded SID is simply that the pattern speed $\Omega_p \equiv \omega/m$ be positive, i.e., that the wave rotates in a prograde sense relative to the (differential) rotation of the disk. In what follows, whenever we discuss spiral disturbances, we adopt the usual convention of choosing m to be a positive integer. Prograde spiral waves then correspond to positive values of ω; retrograde spiral waves to negative values of ω.

If we perturb the disk with a small nonaxisymmetric disturbance (which we denote with a subscript 1), and linearize the fluid equations (5), (6), (7), and (8) about a basic state that is axisymmetric (and denoted by the subscript 0), we get

$$\frac{\partial \Sigma_1}{\partial t} + \frac{1}{\varpi}\frac{\partial}{\partial \varpi}(\varpi \Sigma_0 u_1) + \Omega\frac{\partial \Sigma_1}{\partial \varphi} + \frac{\Sigma_0}{\varpi^2}\frac{\partial j_1}{\partial \varphi} = 0 , \tag{11}$$

$$\frac{\partial u_1}{\partial t} + \Omega\frac{\partial u_1}{\partial \varphi} - 2\Omega\frac{j_1}{\varpi} = -\frac{\partial}{\partial \varpi}\left(\Theta a^2\frac{\Sigma_1}{\Sigma_0} + \epsilon\mathscr{V}_1\right), \tag{12}$$

$$\frac{\partial j_1}{\partial t} + \varpi\frac{\kappa^2}{2\Omega}u_1 + \Omega\frac{\partial j_1}{\partial \varphi} = -\frac{\partial}{\partial \varphi}\left(\Theta a^2\frac{\Sigma_1}{\Sigma_0} + \epsilon\mathscr{V}_1\right), \tag{13}$$

$$\mathscr{V}_1(\varpi, \varphi, t) = -G \oint d\psi \times \int_0^\infty \frac{\Sigma_1(r, \psi, t) r\, dr}{[r^2 + \varpi^2 - 2r\varpi \cos(\psi - \varphi)]^{1/2}} . \tag{14}$$

We look for solutions sinusoidal in t and φ of the form (after taking the real part)

$$\Sigma_1 = S(\varpi)e^{i(\omega t - m\varphi)},$$
$$u_1 = U(\varpi)e^{i(\omega t - m\varphi)},$$
$$j_1 = J(\varpi)e^{i(\omega t - m\varphi)},$$
$$\mathscr{V}_1 = V(\varpi)e^{i(\omega t - m\varphi)}. \quad (15)$$

Substitution of the above expressions into equations (11), (12), (13), and (14) yields

$$i(\omega - m\Omega)S + \frac{1}{\varpi}\frac{d}{d\varpi}(\varpi\Sigma_0 U) - im\Sigma_0\frac{J}{\varpi^2} = 0, \quad (16)$$

$$i(\omega - m\Omega)U - 2\Omega\frac{J}{\varpi} = -\frac{d\Phi}{d\varpi} \quad \text{where}$$

$$\Phi \equiv \Theta a^2\frac{S}{\Sigma_0} + \epsilon V, \quad (17)$$

$$i(\omega - m\Omega)J + \varpi\frac{\kappa^2}{2\Omega}U = im\Phi, \quad (18)$$

$$V(\varpi) = -G\oint d\chi\int_0^\infty \frac{S(r)\cos(m\chi)r\,dr}{[r^2 + \varpi^2 - 2r\varpi\cos\chi]^{1/2}}. \quad (19)$$

We use equations (17) and (18) to eliminate U and J in favor of Φ:

$$U = \frac{i}{(\omega - m\Omega)^2 - \kappa^2}\left[-2\Omega\frac{m}{\varpi} + (\omega - m\Omega)\frac{d}{d\varpi}\right]\Phi, \quad (20)$$

$$\frac{J}{\varpi} = \frac{1}{(\omega - m\Omega)^2 - \kappa^2}\left[(\omega - m\Omega)\frac{m}{\varpi} - \frac{\kappa^2}{2\Omega}\frac{d}{d\varpi}\right]\Phi. \quad (21)$$

The substitution of the above expressions for U and J/ϖ into equation (16) now yields a second-order integro-differential equation first written down without approximation by Lin & Lau (1979):

$$(\omega - m\Omega)S + \frac{1}{\varpi}\frac{d}{d\varpi}\left\{\frac{\varpi\Sigma_0}{(\omega - m\Omega)^2 - \kappa^2}\left[-2\Omega\frac{m}{\varpi} + (\omega - m\Omega)\frac{d}{d\varpi}\right]\Phi\right\} - \frac{m\Sigma_0}{\varpi[(\omega - m\Omega)^2 - \kappa^2]}\left[(\omega - m\Omega)\frac{m}{\varpi} - \frac{\kappa^2}{2\Omega}\frac{d}{d\varpi}\right]\Phi = 0. \quad (22)$$

3. ALIGNED DISTURBANCES

We examine the SID for zero-frequency normal modes by substituting the basic equilibrium state (eq. [9]), setting $\omega = 0$ and separating out a common factor of m. With these substitutions and a little algebra, equation (22) becomes

$$m\left[-S + \frac{1}{D^2(m^2 - 2)}\left(\frac{m^2}{\varpi} - 2\frac{d}{d\varpi} - \varpi\frac{d^2}{d\varpi^2}\right)\times\left(\varpi S + \frac{1 + D^2}{2\pi G}V\right)\right] = 0. \quad (23)$$

Equation (23) is to be solved simultaneously with equation (19).

The $m = 0$ root associated with equation (23) has the following interpretation. With ω and m equal to zero, equations (16) and (18) are satisfied if $U = 0$. The rest of the solution then takes the form $S = K_1/\varpi$, $J = K_2\varpi$, $V = K_3\ln\varpi$, where the ratios of the constants, K_2/K_1 and K_3/K_1, with K_1 small but otherwise arbitrary, are chosen so that equations (17) and (19) can be satisfied. But such a "solution" represents a mere rescaling of one axisymmetric equilibrium to a neighboring axisymmetric equilibrium, which is an allowable but uninteresting perturbation. Henceforth, we focus on the interesting part of equation (23).

The formulation given above is based on an Eulerian description. An alternative formulation for treating linearized perturbations is to consider Lagrangian displacements $\boldsymbol{r}_1$ of fluid elements from their unperturbed positions $\boldsymbol{r}_0$. A picturesque way to describe the collective effect of all such perturbations is to envisage fluid elements displaced from circular orbits to noncircular trajectories that possess m-fold angular symmetry (see, e.g., Kalnajs 1973). When the positions of the maximum inward and outward excursions of fluid elements at different equilibrium radii ϖ_0 line up in azimuth, we call such disturbances *aligned*. When they suffer a systematic shift in angle from one radial position to the next, we call such disturbances unaligned or *spiral*.

In power-law disks, it is clearly interesting to consider aligned perturbations that have the same power-law dependences as the equilibrium state. Such disturbances then have no radial oscillations, and they cannot plausibly be regarded to propagate in the radial direction. Indeed, it is the lack of such propagation that allows the disturbances to be scale-free in the same way that the equilibrium is scale-free. In contrast, spiral disturbances do propagate radially (Toomre 1969). As Syer & Tremaine (1996) were the first to comment, the need for such spiral disturbances to conserve wave angular-momentum and wave energy as they propagate (Shu 1970; Goldreich & Tremaine 1979) implies that they cannot generally have the same self-similarity as their power-law equilibrium states (see also Lee & Goodman 1999 and below).

The above discussion informs us that self-similar aligned (nonspiral) perturbations in a SID must take the form

$$\varpi S(\varpi) = \text{const}, \quad \text{and} \quad V(\varpi) = -\frac{2\pi G}{|m|}\varpi S = \text{const}. \quad (24)$$

The second of the relations in equation (24) is required so that we satisfy Poisson's equation (19) (see proof in forthcoming Paper II).

The substitution of equation (24) into equation (23) now yields the condition for marginal stability as

$$-1 + \frac{m^2}{D^2(m^2 - 2)}\left[1 - \frac{1}{|m|}(1 + D^2)\right] = 0. \quad (25)$$

This equation may be rearranged and factored into the requirement

$$(|m| - 1)[D^2(|m| + 2) - |m|] = 0, \quad (26)$$

which can be satisfied for any D^2 if $|m| = 1$, and which demands

$$D^2 = \frac{|m|}{|m| + 2} \quad \text{for} \quad |m| \geq 2. \quad (27)$$

3.1. *Differences between Aligned and Spiral Instabilities*

In Paper II (forthcoming), we shall rederive the conditions of equation (27) by other means, and we shall demonstrate that they correspond to the onset of bifurcations of axisymmetric SIDs to nonaxisymmetrical SIDs that are more centrally condensed when the rotation rate D is systematically raised to ever higher values. In § 5, we find that the conditions for the onset of such aligned bifurcations are uniformly less stringent than for the onset of swing-amplification of their spiral counterparts for the same value of $|m|$ (see Table 1). For example, barlike deformations with $m = 2$ in equation (27) can occur once $D^2 \geq \frac{1}{2}$ (i.e., once the square of the rotational velocity exceeds half of the square of the magnetosonic speed). This condition is slightly less stringent than the requirement $D^2 \geq 0.5368$ for the amplification of spirals with $m = 2$ arms, which is the most unstable of all small-amplitude spiral waves propagating in an axisymmetric SID (see §§ 4 and 5).

Even more surprisingly, aligned eccentric $m = 1$ displacements can occur that are not a trivial translation of the entire disk for any value of $D^2 \geq 0$. Zang (1976) and Toomre (1977) worried, in a time-dependent spiral context, that such $m = 1$ perturbations, which move the center of the disk with respect to its surroundings, would cause a growing displacement of the center of mass of the system and might therefore be spurious artifacts of the analytical method. We believe that this is not the case since angular momentum can be radiated to infinity in an unbounded SID. In any case, we have no doubts that *aligned* eccentric displacements are possible alternative states of *equilibria* for extended SIDs that bypass the question of whether or not such states can be reached by continuous evolution in time from an initially axisymmetric state (see Paper II, forthcoming).

The instability mechanisms for aligned and spiral instabilities are fundamentally different. Aligned deformations occur, if there is a dynamic or dissipative route to tap the energy difference, because it is energetically favorable for a rapidly enough rotating system to become nonaxisymmetric, rather than remain axisymmetric, by pulling some of the material closer to the origin at the expense of sticking some of it farther out—the combination is needed to conserve total angular momentum. Wave propagation plays no role in this analysis. In contrast, wave propagation plays a crucial role in spiral instabilities—in particular, in considerations of wave amplification across the corotation circle (see § 5).

The above comments make clear that the homology or truly scale-free transformations belong to a fundamentally different class of disturbances than do spiral modes. (Toomre [1981] expresses a variant of this point of view when he speculates that "edge modes" may be analogs of the bifurcation of Maclaurin spheroids to Jacobi ellipsoids.) Among other qualities, the aligned disturbances possess rigorous extensions to the finite-amplitude regime (Paper II, forthcoming). Just as Poincaré (1885) and Liapunov (1905) thought incompressible ellipsoidal figures of equilibrium appropriate candidates for fissioning via pear-shaped equilibria into binary stars (see also Jeans 1928), so it is a worthwhile speculation that nonaxisymmetric, magnetized, compressible, and perhaps truncated SIDs might prove to be promising candidates for fragmentation into binary and multiple star systems. The fission hypothesis foundered on numerical simulations of compressible bodies (e.g., Durisen et al. 1986) that showed transport of mass and angular momentum via bars and spiral density waves stabilized unstable configurations before they could fission. The same fate might befall the nonaxisymmetric SIDs, as is suggested by the work of Laughlin & Bodenheimer (1994), but an ace in the hole, rapid magnetic flux loss, may yet save the day (see Paper III, forthcoming).

TABLE 1

CRITICAL VALUE OF D^2 FOR ALIGNED AND SPIRAL DISTURBANCES

$\lvert m\rvert$	Aligned	Spiral
1	0.0000	0.7560
2	0.5000	0.5368
3	0.6000	0.6180
4	0.6667	0.6776
5	0.7143	0.7217
... .	...	...
∞	1.0000	1.0000

3.2. *Secular Barlike Instability*

In galactic dynamics, Ostriker & Peebles (1973; see also Hohl 1971; Miller 1971; Kalnajs 1972) postulated the existence of massive dark halos to suppress strong $m = 2$ barlike instabilities relative to spiral instabilities. In star-forming clouds, one does not have such freedom, and one has to contend with the real existence and threat of secular instabilities posed by nonaxisymmetric aligned disturbances of symmetry $m = 1$ and $m = 2$. In the case $m = 2$, these modes are analogs of the homologous perturbations that convert Maclaurin spheroids to Jacobi or Dedekind ellipsoids, etc. (see Chandrasekhar 1969). (The pattern speed of Dedekind ellipsoids is zero, so they are the better analogy.)

The criterion for (secular and dynamic) barlike instability is often expressed in terms of the ratio of the kinetic energy of rotation $\mathcal{T}$ to the absolute value of the gravitational potential energy $\mathcal{W}$ (cf. Ostriker & Peebles 1973). This is difficult to do for an unbounded SID since both quantities are formally infinite and a limit process is needed. We begin by redefining $\mathcal{W}$ so that it equals the gravitational work integral for $\varpi \leq R$, which is equal to the usual gravitational potential energy in the scalar virial theorem if there is no matter outside R. To obtain the scalar virial theorem in the present case, we multiply the radial force-balance equation of the equilibrium state,

$$-\Sigma_0 \varpi \Omega^2 = -\frac{d}{d\varpi}(\Theta \Pi_g) - \epsilon \Sigma_0 \frac{d\mathcal{V}_0}{d\varpi}, \tag{28}$$

by $2\pi\varpi^2\, d\varpi$ and integrate from 0 to R. An integration in parts of the pressure term and a little algebra then yields

$$2(\mathcal{T} + \Theta \mathcal{U}) + \epsilon \mathcal{W} = 2\pi R^2 \Theta \Pi_g(R), \tag{29}$$

where

$$\mathcal{T} \equiv \int_0^R \tfrac{1}{2}\Sigma_0(\varpi\Omega)^2 2\pi\varpi\, d\varpi,$$

$$\mathcal{U} \equiv \int_0^R \Pi_g\, 2\pi\varpi\, d\varpi,$$

$$\mathcal{W} \equiv -\int_0^R \varpi\Sigma_0 \frac{d\mathcal{V}_0}{d\varpi}\, 2\pi\varpi\, d\varpi. \tag{30}$$

To derive equation (29), we have assumed that $\varpi^2\Pi_g \to 0$ in the limit $\varpi \to 0$, which is valid for a SID. Notice also that the thermal energy $\mathscr{U}$ is obtained as the integral of Π_g rather than $(3/2)\Pi_g$ because a thin disk effectively has only two translational degrees of freedom rather than three. The surface term on the right-hand side of equation (29) makes a nonvanishing contribution $\propto R$ for an infinite SID where $\Pi_g = a^2\Sigma_0 \propto 1/\varpi$ for ϖ larger than R.

We use equation (9) to compute the various integrals as

$$\mathscr{T} = \frac{\Theta^2 a^4 D^2(1+D^2)}{2\epsilon G} R ,$$

$$\mathscr{U} = \frac{\Theta a^4(1+D^2)}{\epsilon G} R ,$$

$$\mathscr{W} = -\frac{\Theta^2 a^4(1+D^2)^2}{\epsilon^2 G} R . \quad (31)$$

Substitution of equation (31) into the scalar virial theorem, with the surface term evaluated as $2\pi R^2\Theta\Pi_g(R) = [\Theta^2 a^4 (1+D^2)/\epsilon G]R$, shows that equation (29) is satisfied identically, as it must be. The ratio $\mathscr{T}/|\epsilon\mathscr{W}|$, where $\epsilon\mathscr{W}$ is the *equivalent* gravitational-work integral including the effects of magnetic tension, can now be calculated as the simple expression

$$\frac{\mathscr{T}}{\epsilon|\mathscr{W}|} = \frac{D^2}{2(1+D^2)} . \quad (32)$$

Notice that $\mathscr{T}/\epsilon|\mathscr{W}|$ attains the naively expected ratio $\frac{1}{2}$ in the limit that the disk becomes rotationally supported, $D^2 \to \infty$, because pressure terms in the virial theorem, in the interior or at the surface, become unimportant in this limit.

For $D^2 = \frac{1}{2}$, the critical value for barlike bifurcation in an infinite SID, $\mathscr{T}/\epsilon|\mathscr{W}| = \frac{1}{6} = 0.1667$, which is somewhat larger than other estimates extant in the literature for secular instability to barlike perturbations in unmagnetized configurations (e.g., Ostriker & Peebles 1973; Hunter 1977; Bardeen et al. 1977). We believe that this discrepancy arises because, in some crude sense, the gravitational work integral over finite R underestimates (in absolute value) the self-gravitational energy of that region *plus* the potential energy of matter inside that region caused by the gravity of the matter outside that region. In any case, the discrepancy is slight for most applications of astronomical importance.

In §§ 5.7 and 7, we will argue that $D^2 = 0.5368$ defines the onset of a dynamical barred-spiral instability in an appropriately modified SID. Such a value corresponds to $\mathscr{T}/\epsilon|\mathscr{W}| = 0.1743$, which is coincidentally close to the secular value 0.1667.

4. SPIRAL DENSITY WAVES

To examine the properties of spiral density waves with zero pattern speed, we follow Syer & Tremaine (1996) and Lynden-Bell & Lemos (1993) and note that equations (19) and (23) allow scale-free solutions of the form

$$S(\varpi) = s\varpi^{-3/2}e^{i\alpha\ln\varpi} , \quad V(\varpi) = v\varpi^{-1/2}e^{i\alpha\ln\varpi} , \quad (33)$$

with s, v, α equal to constants. The substitution of equation (33) into equation (19) yields

$$v = -2\pi G\mathscr{N}_m(\alpha)s , \quad \text{where} \quad \mathscr{N}_m(\alpha) \equiv K(\alpha, m) , \quad (34)$$

with $K(\alpha, m)$ being the function even in m and α being defined by Table 1 of Kalnajs (1971) and shown for $m = 0$–2 in Figure 1. Kalnajs (1971) also has derived a recursion relation for fixed α that is useful for numerical computations,

$$\mathscr{N}_{m+1}(\alpha)\mathscr{N}_m(\alpha) = [(m+\tfrac{1}{2})^2 + \alpha^2]^{-1} . \quad (35)$$

Thus, for $\alpha = 0$, $\mathscr{N}_m(\alpha) = 4.376879, 0.913893, 0.486320, \ldots$, when $m = 0, 1, 2, \ldots$. For large m or large α, $\mathscr{N}_m(\alpha) \approx (m^2+\alpha^2)^{-1/2}$; indeed, this asymptotic approximation is accurate to better than 3% for all α and m once $|m| \geq 2$. To this level of accuracy, we may extend the validity of the formula down to even $|m| = 1$ by using

$$\mathscr{N}_m(\alpha) \approx (m^2+\alpha^2+\tfrac{1}{4})^{-1/2} . \quad (36)$$

The substitution of equation (33) into equation (23) now yields for the curve of marginal stability, either $m = 0$ (transformations to neighboring equilibria) or

$$-1 + \frac{1}{D^2(m^2-2)}\left(m^2+\alpha^2+\frac{1}{4}\right) \times [1-(1+D^2)\mathscr{N}_m(\alpha)] = 0 , \quad (37)$$

which indicates that the disk will support discrete nonrotating logarithmic spirals that depend on the number of arms m and the circular speed D. A remarkable cancellation of terms proportional to $i\alpha$ is seemingly needed to obtain equation (37), which is even in α and m. This cancellation is readily understood when one recalls that the antispiral theorem (Lynden-Bell & Ostriker 1967) guarantees the complete equivalence of leading and trailing spirals (positive and negative values of α) when there is no time-dependence in the problem (associated, for example, with corotation resonances or with wave growth or damping).

Equation (37) may be written in the physically more suggestive form,

$$m^2D^2 = 2D^2 + (m^2+\alpha^2+\tfrac{1}{4}) - (1+D^2)(m^2+\alpha^2+\tfrac{1}{4})\mathscr{N}_m(\alpha) , \quad (38)$$

which may be advantageously compared with the standard WKBJ relationship for spiral density waves (Lin & Shu 1966, 1968):

$$(\omega - m\Omega)^2 = \kappa^2 + k^2\Theta a^2 - 2\pi\epsilon G|k|\Sigma_0 . \quad (39)$$

We recognize the coefficients in equation (38) as the dimensionless form of those in equation (39), when ω is set equal to zero and the substitution of equation (9) is made in the latter. The analogy is completed by utilizing the approximation of equation (36) in equation (38) and identifying the effective wavenumber k as

$$k = \alpha_e\varpi^{-1} \quad \text{where} \quad |\alpha_e| \equiv (m^2+\alpha^2+\tfrac{1}{4})^{1/2} . \quad (40)$$

Since in the simplest form of the WKBJ approximation, k is taken to be the radial wavenumber α/ϖ, we refer to the identification in equation (40) as the *extended* WKBJ approximation.

If we want to attribute a fraction $(1-F)$ of the gravity in the equilibrium state to an axisymmetric dark halo that is unresponsive to spiral perturbations, we should replace the term $(1+D^2)$ that appears in equations (9) and (38), and everywhere else it appears below, by $F(1+D^2)$. We call disks partial SIDs if $0 < F < 1$, in contrast with full SIDs for which $F = 1$. Regarded as an equation for $|\alpha|$, given

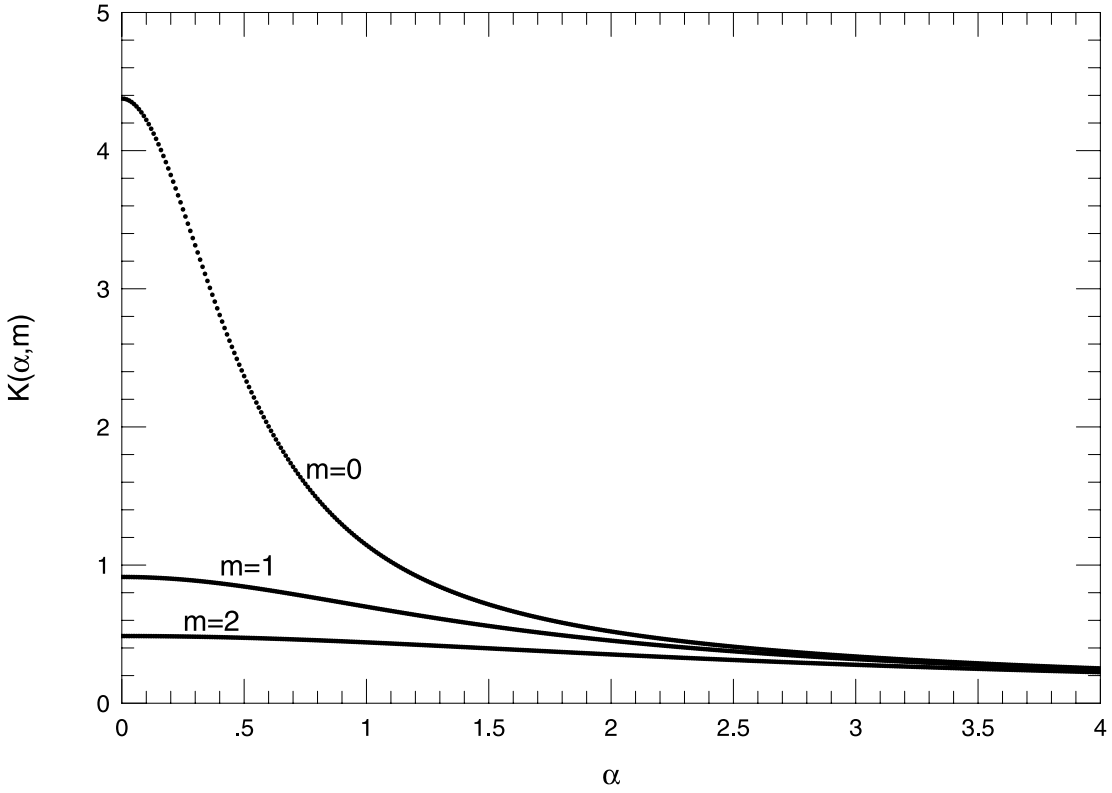

FIG. 1.—Plot of the Kalnajs function $K(\alpha, m) \equiv \mathcal{N}_m(\alpha)$ as a function of the dimensionless radial wavenumber α in a logarithmic spiral for $m = 0$–2. The quantity m/α is the tangent of the angle of the m-armed spiral pattern with respect to the circular direction; i.e., large values of $|\alpha|$ correspond to tightly wound spirals.

$m \neq 0$ and ω, the dispersion relation in equation (39) applied to a partial SID can yield two meaningful solutions for $|\alpha|$, corresponding to long and short spiral waves, in circumstances where the full SID would only have one solution (short spiral waves, which can then be leading or trailing depending on the sign of α). An extreme example is Saturn's rings (with a near-Keplerian rotation curve, maintained mostly by the planet), where the dominant density waves of interest are long waves, not short ones (see Goldreich & Tremaine 1978; Shu 1984). In situations where both long and short spiral waves can be present—which include protoplanetary disks—the variety of wave-wave interactions that can take place is richer than those discussed in § 5 (see Adams, Ruden, & Shu 1989; Shu et al. 1990).

By analogy with the usual physical interpretation for the WKBJ dispersion relationship, we regard the exact equation (38) to yield the required value of D^2 to marginally stabilize a disturbance of given dimensionless and real radial wavenumber α. The interpretation in terms of growing normal modes turns out to be straightforward only in the case $m = 0$ (see below); for $m \neq 0$, "marginal stability" really means whether *propagating* spiral density waves can be swing-amplified (see § 5). With this understanding, the "curve of marginal stability" becomes

$$D^2 = \frac{(m^2 + \alpha^2 + \frac{1}{4})[\mathcal{N}_m(\alpha) - 1]}{2 - m^2 - (m^2 + \alpha^2 + \frac{1}{4})\mathcal{N}_m(\alpha)} . \tag{41}$$

4.1. *Axisymmetric Instability*

Consider first the noncontroversial case of axisymmetric instability (Toomre 1964; see also Schmitz 1986, 1988, 1990; Lemos, Kalnajs, & Lynden-Bell 1991). For $m = 0$, D^2 as given by equation (41) starts off at 0.9320 at $|\alpha| = 0$, and it declines to zero for $|\alpha| = 1.120$ [where $\mathcal{N}_0(\alpha) = 1$]. There is then a gap in $|\alpha|$ where no positive value for D^2 can be found before a second branch of the curve of marginal stability starts at $|\alpha| = 1.759$ [where $(\alpha^2 + \frac{1}{4})\mathcal{N}_0(\alpha) = 2$] with D^2 appearing at $+\infty$. This second branch reaches a minimum at $|\alpha| = 3.056$ where $D^2 = 5.410$ before D^2 starts increasing again roughly linearly with $|\alpha|$ at large $|\alpha|$. These curves of marginal stability are shown in Figure 2. For comparison, we note that the effective Toomre Q parameter for our problem reads

$$Q \equiv \frac{\Theta^{1/2} a\kappa}{\pi\epsilon G\Sigma_0} = 2\sqrt{2}\,\frac{D}{1 + D^2} , \tag{42}$$

and Q equals 1.413 when $D^2 = 0.9320$, and it equals 1.026 when $D^2 = 5.410$. For a full SID, notice that Q has a maximum value of $\sqrt{2}$ at $D = 1$, and that Q falls to unity on either side of this maximum at $D = \sqrt{2} \pm 1$, or at $D^2 = 0.1716$ and $D^2 = 5.828$.

The physical interpretation of the curves of marginal stability is as follows. On the first branch of the D^2 versus $|\alpha|$ curve, the SID is unstable to axisymmetric collapse for $D^2 < 0.9320$ ($Q < 1.413$) for the range of wavenumbers $|\alpha|$ between the intersection of the line $D^2 = \text{const}$ corresponding to the equilibrium model and the curve. (For an explicit calculation of this collapse in the special case $D = 0$, see Li & Shu 1997.) Above the curve, with $D^2 > 0.9320$, the SID rotates too fast to be unstable with respect to axisymmetric collapse for any value of the logarithmically sinusoidal wavenumber α. In particular, stable "breathing modes" are possible at $\alpha = 0$ for $D^2 > 0.9320$ (see Lemos et al. 1991 for the general criteria in the case of power-law disks).

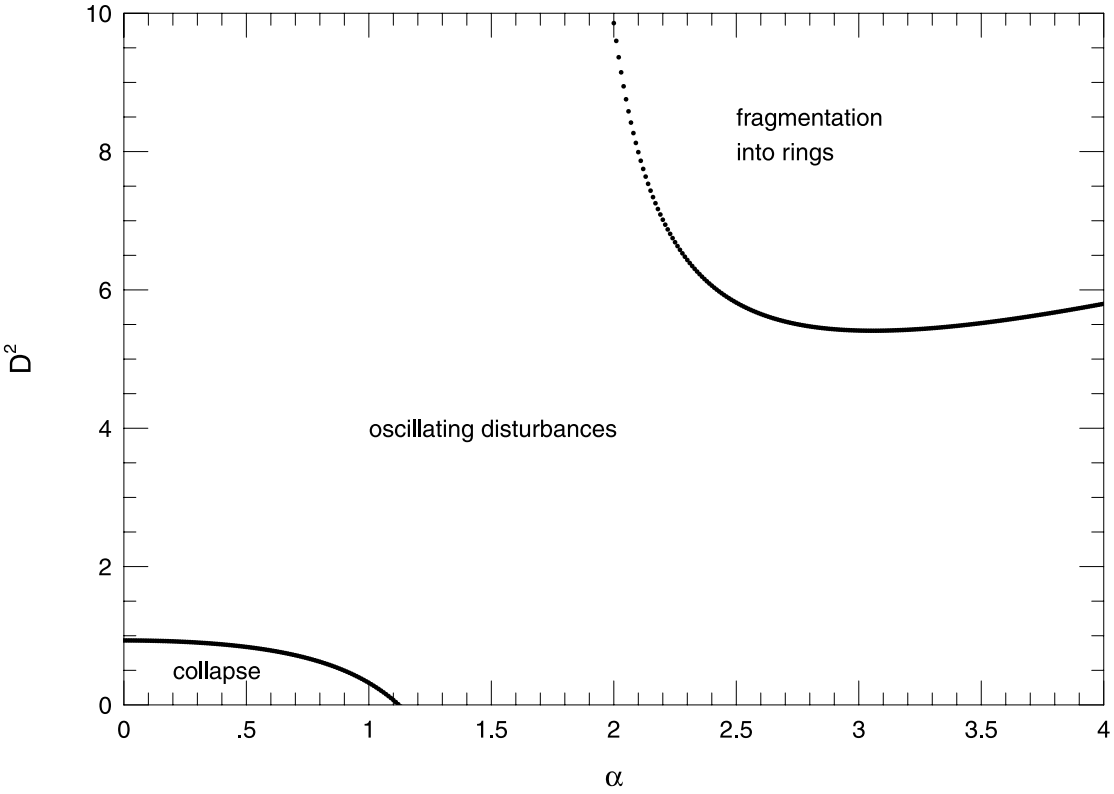

FIG. 2.—Curves of marginal stability for axisymmetric (ringlike) disturbances in a SID

The SID does not remain axisymmetrically stable, however, for arbitrarily higher rotation rates $|D|$. When D^2 exceeds 5.410, intersections of horizontal lines of $D^2 = \text{const}$ can occur with the second branch of the curve of marginal stability. For $D^2 > 5.410$ ($Q < 1.026$), the SID becomes unstable to ringlike fragmentation for the range of wavenumbers α that lie between the two points of intersection between the horizontal line and the second branch of the marginal-stability curve. Notice that Toomre's approximate criterion, $Q < 1$ for axisymmetric instability, yields a more accurate estimate for fragmentation, $Q < 1.026$, than it does for collapse, $Q < 1.413$, because the required range of $|\alpha|$ is larger in the former case than in the latter and, therefore, more closely satisfies the formal requirement $|\alpha| \gg 1$ needed for the WKBJ asymptotic analysis to apply.

4.2. *Spiral Perturbations*

Consider next the case when $m \neq 0$. Setting $\omega = 0$ and adopting the spiral waveform of equation (33) has no simple interpretation in terms of some limit of a growing spiral disturbance that also propagates since the principle of exchange of stabilities (ω going from real to imaginary through zero) does not apply to nonaxisymmetric instabilities. Nevertheless, there is a sense (to be defined below) in which we may regard equation (37) as yielding the curves of marginal stability for the cases $m = 1, 2, 3, \ldots$.

The locus of marginal stability can also be computed approximately by applying the extended WKBJ approximation recommended in equations (36) and (40) to equation (41). This yields

$$D^2 \approx \frac{|\alpha_e|(|\alpha_e| - 1)}{|\alpha_e| + m^2 - 2}. \tag{43}$$

Figure 3 shows that the approximation is very good for $m \geq 2$. The accurate criterion implies that the SID can support one-armed zero-frequency spirals when $D^2 > 0.7560$, before it becomes stable to axisymmetric collapse, $D^2 > 0.9320$, and before it becomes unstable to ringlike fragmentation, $D^2 > 5.410$. Similarly, for $m = 2$, we find that the SID can support two-armed spirals when $D^2 > 0.5368$, corresponding to a critical $Q = 1.348$. The critical value that D^2 must exceed for various aligned and spiral instabilities is summarized in Table 1. For D^2 marginally above the critical value 0.5368, the full SID is less hospitable for spirals with $|m| \neq 2$. The reversal of sign in the term $m^2 - 2 \propto m^2\Omega^2 - \kappa^2$ that measures the frequency distance from Lindblad resonance implies that it becomes progressively harder for the disk to respond in phase with its own spiral-wave forcing when $|m|$ exceeds 2. However, the full SID can be unstable to many spiral instabilities even when it is stable to axisymmetric ones (of the density-wave type). This conclusion also holds in non–scale-free disks, where wave feedback plus wave amplification across the corotation circle can give globally unstable spiral modes. But there, if we are dealing with stellar disks rather than gaseous ones, wave absorption at inner Lindblad resonances often effectively suppresses such feedback and amplification loops for $|m| = 3, 4, 5, \ldots$.

In a gaseous disk, Figure 3 shows that $|m| = 2$ disturbances require the smallest value of D^2 to suppress zero-frequency spirals with $\alpha \approx 0$. Indeed, when $D^2 < 0.5386$, the full SID cannot support zero-frequency spiral waves of any m or α. As we shall see in § 5, it can still support nonlogarithmic spiral waves of nonzero wave frequency ω in the outer disk, but the condition $D^2 < 0.5386$ in a full SID leads to the stabilization of all such waves against overreflection near the corotation circle. More generally, we may

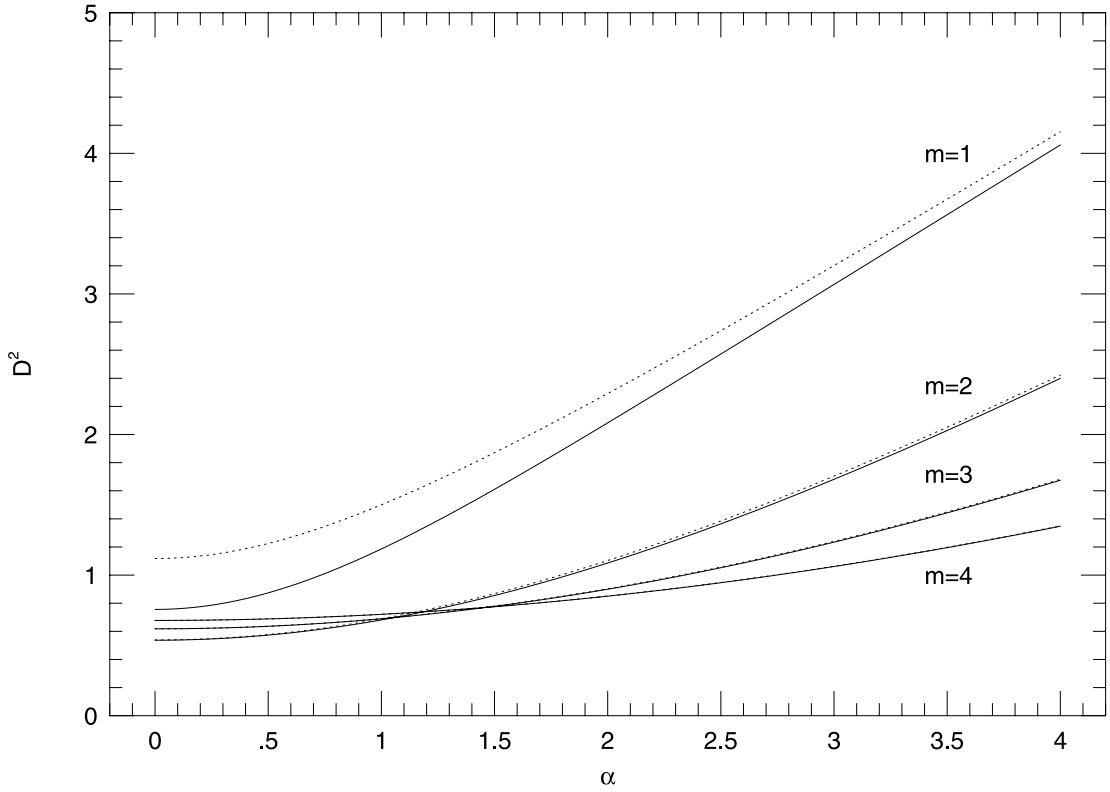

FIG. 3.—Self-consistency curves for nonrotating spiral disturbances of $m = 1$–4 in a SID. Solid lines show the relationship between dimensionless squared rotation-rate D^2 and dimensionless radial wavenumber α for a logarithmic spiral obtained with the exact equation (41); dashed lines show the same relationship obtained with the approximate equation (43). For $m \geq 2$, these curves are indistinguishable at the resolution of the plot.

state that for amplification of m-armed spiral disturbances to be possible, D^2 must exceed the critical value obtained by setting α equal to zero on the right-hand side of equation (41). This criterion for swing amplification is identical to the one that Goodman & Evans (1999, eq. [76]) independently obtain for the onset of instability in their conceptually and technically different treatment of normal modes in a SID.

5. WAVE AMPLIFICATION ACROSS THE COROTATION RADIUS

The introduction of a spiral disturbance rotating with a pattern speed ω/m defines a length scale, given by a corotation radius (CR), into the problem. For ω real and positive, we can simplify the analysis by defining a renormalized radial coordinate ξ by

$$\xi \equiv \omega\varpi/D\Theta^{1/2}a \, . \tag{44}$$

With this definition, corotation occurs at $\xi = m$, if we also adopt the usual convention that $m \geq 0$. The outer and inner Lindblad resonances (OLR and ILR) occur at

$$\xi = L_{\pm}, \quad \text{where} \quad L_{\pm} = m \pm \sqrt{2} \, . \tag{45}$$

OLRs exist for any $m \geq 1$, but ILRs exist inside the disk only for $m \geq 2$. For later reference, we notice the WKBJ Q-barriers begin and end at

$$\frac{(\xi - m)^2}{2} = 1 - \frac{1}{Q^2} \, , \quad \text{i.e.,} \quad \xi = L_{\pm} \mp \Delta \, , \tag{46}$$

where

$$\Delta \equiv \sqrt{2}\left[1 - \left(1 - \frac{1}{Q^2}\right)^{1/2}\right] \geq 0 \quad \text{for} \quad Q \geq 1 \, . \tag{47}$$

However, the location of the inner edge of the Q-barrier generally is not calculated accurately by the simple WKBJ theory (the extended version, with $|\alpha_e|$ defined in eq. [40], does better). For that purpose, the criterion in equation (41) is the preferred method for assessing the existence of propagating density waves (see below).

With the introduction of the above notation and the adoption of the equilibrium state (full SID) given by equation (9), equations (22) and (19) become

$$D^2(\xi - m)S + \frac{d}{d\xi}\left\{\frac{1}{[(\xi - m)^2 - 2]}\left[-2m + \xi(\xi - m)\frac{d}{d\xi}\right] \times [\xi S + (1 + D^2)W]\right\} - \frac{m}{[(\xi - m)^2 - 2]}\left[(\xi - m)\frac{m}{\xi} - \frac{d}{d\xi}\right] \times [\xi S + (1 + D^2)W] = 0 \, , \tag{48}$$

$$W(\xi) = -\frac{1}{2\pi}\oint \cos(m\psi)d\psi \times \int_0^{\infty} \frac{S(\eta)\eta \, d\eta}{(\eta^2 + \xi^2 - 2\xi\eta\cos\psi)^{1/2}} \, . \tag{49}$$

Notice that the recurring term $[(\xi - m)^2 - 2]$ above factors into $(\xi - L_+)(\xi - L_-)$.

5.1. *Governing Equations for Swing Amplification in a Full SID*

At this point, we introduce the variables

$$X(\xi) \equiv \xi S(\xi), \quad Y(\xi) = \xi S(\xi) + (1 + D^2)W(\xi) \, , \tag{50}$$

and rewrite equations (48) and (49) as coupled integro-

differential equations:

$$\xi^2 \frac{d^2Y}{d\xi^2} + \xi(2 - \mathscr{D}) \frac{dY}{d\xi} + \left(\frac{2m\mathscr{D}}{\xi - m} - m^2\right)Y + D^2[(\xi - m)^2 - 2]X = 0 , \quad (51)$$

$$Y(\xi) = X(\xi) - \frac{(1 + D^2)}{2\pi} \oint \cos(m\psi) d\psi \times \int_0^\infty \frac{X(\eta) d\eta}{(\eta^2 + \xi^2 - 2\xi\eta \cos\psi)^{1/2}} . \quad (52)$$

In the above equations, we have denoted

$$\mathscr{D} \equiv \frac{d \ln [(\xi - m)^2 - 2]}{d \ln \xi} = \frac{2\xi(\xi - m)}{[(\xi - m)^2 - 2]} \quad (53)$$

when we carry out the indicated differentiation.

The term $2m\mathscr{D}/(\xi - m)$ in equation (51) gives what is known as the corotation resonance. In our problem, the corotation point does not appear explicitly in the form of a resonant denominator because the coefficient $\mathscr{D}$ is itself proportional to the combination $\xi - m$. In other words, the term that Lin & Lau (1979) denote as B_s is identically equal to zero, or, in the terminologies of Papaloizou & Lin (1989) or Goodman & Evans (1999), the *vortensity* or *potential vorticity*, has zero gradient in our model disk. For given X, however, $\xi = 0$, L_+, and L_- are regular singular points of the ordinary differential equation (51) for Y.

5.2. *Asymptotic Forms*

For $\xi \ll m$, $\mathscr{D} \approx -2m\xi/(m^2 - 2)$ can be ignored in equation (51), whereas equation (52) has the approximate scale-free solution near the origin:

$$X \approx c_\pm \xi^{-1/2} e^{\pm i|\alpha| \ln \xi},$$
$$Y = [1 - (1 + D^2)\mathscr{N}_m(\alpha)]X , \quad (54)$$

with the constant $|\alpha|$ to be determined and $c_\pm$ are arbitrary complex constants in the linearized theory. Substitution of the above into the dominant terms in equation (51) yields the requirement for ξ near zero:

$$-(m^2 + \alpha^2 + \tfrac{1}{4})[1 - (1 + D^2)\mathscr{N}_m(\alpha)] + D^2[m^2 - 2] = 0 . \quad (55)$$

This equation is equivalent to the criterion in equation (37), except that we now identify $|\alpha|$ to be the required eigenvalue for given D^2 and nonzero ω. That such an $|\alpha|$ exists is guaranteed if D^2 was originally chosen so that equation (37) has a solution for some real α.

For $\xi \gg m$, $\mathscr{D} \approx 2$ and the dominant terms in equation (51) read

$$\xi^2\left[\frac{d^2Y}{d\xi^2} + D^2X\right] = 0 ; \quad (56)$$

whereas equation (52) has the approximate solution

$$Y \approx X . \quad (57)$$

Far beyond the OLR, spiral density waves in a gaseous disk have negligible self-gravity and become essentially sound waves with the solution (the extra D comes from defining ξ in eq. [44] in terms of the rotation speed instead of the sound speed)

$$Y \approx X \approx C_\pm e^{\pm iD\xi} , \quad (58)$$

where $C_\pm$ are again arbitrary complex constants.

The solution for the scaled gravitational potential reads

$$(1 + D^2)W = Y - X \approx \frac{\mathscr{C}_\pm}{\xi} \exp\left\{\pm iD\left[\xi + \left(\frac{1 + D^2}{2D} - m\right)\ln\xi\right]\right\}, \quad (59)$$

where $\mathscr{C}_\pm$ are complex constants.

5.3. *Formulation of Overreflection Problem*

The wave-amplification problem can now be formulated as follows. If we restrict the analysis to full SIDs, the dispersion relation in equation (37) supports only a single, short-wavelength branch. We look for solutions that correspond to a single spiral wave (either leading or trailing) in the region near the origin that is asymptotically far from the corotation radius:

$$X = c_+ \xi^{-1/2} e^{+i|\alpha| \ln \xi} \quad \text{or} \quad X = c_- \xi^{-1/2} e^{-i|\alpha| \ln \xi} . \quad (60)$$

The c_+ or c_- solution will attach to a solution in the asymptotic region outside the CR that consists of a superposition of leading and trailing spiral components, C_+ and C_-.

The c_- solution (corresponding to a trailing spiral interior to the corotation radius) has a simple interpretation in terms of swing amplification (see Toomre 1981): a leading spiral C_+ emitted from infinity exterior to the corotation radius impinges on the corotation region and transmits a single trailing wave c_- interior to corotation (after tunneling through the Q-barrier), in the process it overreflects to a trailing spiral wave C_- propagating back to infinity away from corotation. The overreflection factor (in wave energy or angular momentum) is defined by the ratio $|C_-|^2/|C_+|^2$. As a shorthand notation, we denote the *physical process* as $C_+ \to C_-$, c_- and the *mathematical procedure* for finding this solution as $c_- : C_+, C_-$.

The other solution $c_+ : C_+, C_-$ has no such simple interpretation. The outer-cavity waves C_+ and C_- might represent emitted and reflected disturbances, but the (short) leading wave c_+ propagates toward CR, not away from it. Therefore, c_+ cannot be interpreted as a transmitted wave. However, we may take a linear combination of the two elementary solutions, $c_+ : C_+^+, C_-^+$ and N times $c_- : C_+^-, C_-^-$ to cancel out the traveling wave $C_+ = C_+^+ + NC_+^- \equiv 0$ exterior to CR. The last equation defines the needed complex number $N = -C_+^+/C_+^-$. The combination that gives rise to a single trailing wave exterior to CR (i.e., with $C_+ = 0$ and $C_- = C_-^+ + NC_-^- \neq 0$) then has an equally simple interpretation: $c_+ \to c_-, C_-$. A leading spiral c_+ is emitted from the origin and travels toward CR. At the Q-barrier, c_+ overreflects into a trailing spiral wave c_- (with the overreflection factor equal to $|c_-|^2/|c_+|^2$) that propagates back toward the origin, in the process transmitting, by tunneling under the Q-barrier, a single trailing wave C_- that radiates to infinity.

5.4. *Numerical Procedure*

Our numerical procedure for finding solutions is a variant of the matrix technique that has been repeatedly applied to equation (22) (see, e.g., Pannatoni 1979; Adams, Ruden, & Shu 1989; or Laughlin 1994). In the present implementation, the integrodifferential governing equations (51) and (52) are cast in finite-difference form on a $n = 1024$ point logarithmically spaced grid extending from $\xi_{\rm in} = 0.0005$ to $\xi_{\rm out} = 40.0$. The Poisson kernels in equation (52)

are integrated from ξ_{in} to ξ_{out} using an iterative quadrature scheme. Divergences at the disk edges are avoided by using a squared softening length, $\delta^2 = 0.0001$, in the denominator of equation (52). The governing equations are then expressed as a homogeneous matrix equation, with the coefficient matrix being full of elements because the Poisson operator connects all grid points to each other.

The second-order nature of the ordinary differential equation (23) means that we are free to specify one complex amplitude at each boundary. At the inner boundary, we set $X = 30 + 0i$ (multiplication by a complex constant of unit modulus merely rotates the pattern through a fixed angle), and at the outer boundary we get two linearly independent solutions by setting in turn $X = 1 + 0i$ and $X = 0 + 1i$. With the boundary conditions at the edges incorporated, we have an inhomogeneous 1022 point matrix equation, which can be inverted to solve for the eigenvector itself [discretized values of $X(\xi)$]. Once the two linearly independent solutions are found, we are free to superimpose them in the manner previously discussed to obtain physically interpretable solutions. In particular, our ultimate choice of radiation boundary conditions for the output end implies that we operationally have wave absorbers for the transmitted wave.

We find that the grid truncation and gravity softening modify the values of α and D required to fit the asymptotic solutions in equations (54) and (58), which are valid as $\ln \xi \to \mp\infty$. In practice, then, we measure the "observed" value D' in the outer cavity by taking

$$D' = \frac{2\pi}{\xi_{2\,\max} - \xi_{1\,\max}}, \tag{61}$$

where $\xi_{2\,\max}$ and $\xi_{1\,\max}$ are the radial locations of successive wave crests of the outer asymptotic solution. Alternately, D' can be measured by Fourier analyzing the waveform in the outer cavity.

If we sample the numerically obtained solution vector at two points, ξ_1 and ξ_2 in the outer cavity, and assume that the waveform is given by $X \approx C_\pm e^{\pm iD'\xi}$, then we have

$$X_1(\xi_1) = C_+ e^{iD'\xi_1} + C_- e^{-iD'\xi_1}, \tag{62}$$

$$X_2(\xi_2) = C_+ e^{iD'\xi_2} + C_- e^{-iD'\xi_2}, \tag{63}$$

which can be solved for the complex constants $C_\pm$. An exactly similar analysis yields the value of α' and the complex constants $c_\pm$ required to match the asymptotic waveform obtained by the matrix method in the interior cavity

$$\xi^{1/2} X = c_\pm e^{\pm i|\alpha'| \ln \xi}. \tag{64}$$

After the first linearly independent numerical solution has been fitted to the asymptotic form in the inner and outer cavities of the disk, we then fit the second numerical solution. This allows us to combine the solutions to give a single trailing wave exterior to corotation. The inner asymptotic region of the combined solution can then be analyzed by the method described above to determine the overreflection factor associated with the corotation point.

5.5. *Numerical Results*

In Figure 4, we plot $m = 2$ solutions obtained with boundary values $X(\xi_{in}) = 30$ and $X(\xi_{out}) = i$ for $D = 0.75$, 1.05, 1.35, and 1.65. These solutions approach the asymptotic behavior predicted by equations (54) in the inner region and (58) in the outer region. As D approaches the minimum value $D = \sqrt{0.5368} = 0.7327$ for supporting zero-frequency 2-armed spirals, the wavelength in the inner cavity becomes infinite. A physical interpretation is provided in the next subsection. The result is that a leading wave propagating inward from large radii reflects off the corotation radius without transmitting a signal across the Q-barrier. As D becomes larger, transmission across the Q-barrier becomes increasingly easy, and overreflection becomes large. By $D = 1.65$, it is difficult to discern the evanescent character of the disturbance within the Q-barrier.

Figure 5 shows the overreflection factors obtained from the pure inner cavity process $c_+ \to c_-$, C_-, plotted as a function of the rotation speed D. Because of the disk truncation and gravity softening, our asymptotic solutions are not perfect, and it is not possible to obtain a perfect trailing spiral in the exterior cavity. This is especially true for disks that have powerful corotation amplifiers that generate large overreflection factors. In Figure 5, we plot overreflection factors obtained with two different choices of the fitting points ξ_1 and ξ_2. This gives a better idea of both the error intrinsic to the measurement procedure as well as the actual trend of results. Our inability to obtain high numerical accuracy for the overreflection factor in Figure 5 is related to similar computational problems experienced by Goodman & Evans (1999, Fig. 3) in estimating the related ratio of growth rate to pattern speed in a normal-modes treatment of the SID.

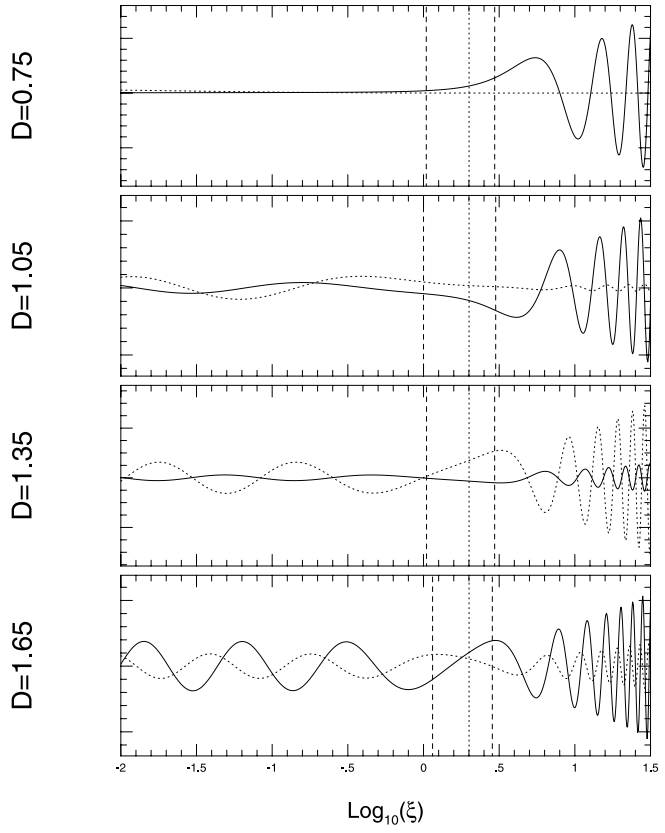

FIG. 4.—Numerical solutions (expressed as $\xi^{1/2}X$ for clarity) to the coupled integrodifferential equations (51) and (52), for $m = 2$, and (*from top to bottom*) $D = 0.75$, 1.05, 1.35, and 1.65. The corotation radius (located at $\xi = 2.0$) is indicated by short-dashed vertical lines. The locations of the Q-barriers, as computed by the approximate equations (46) and (47), surrounding the corotation radius are shown as vertical lines composed of longer dashes. Re $(\xi^{1/2}X)$ is plotted as a solid curve, whereas Im $(\xi^{1/2}X)$ is plotted with a dashed curve.

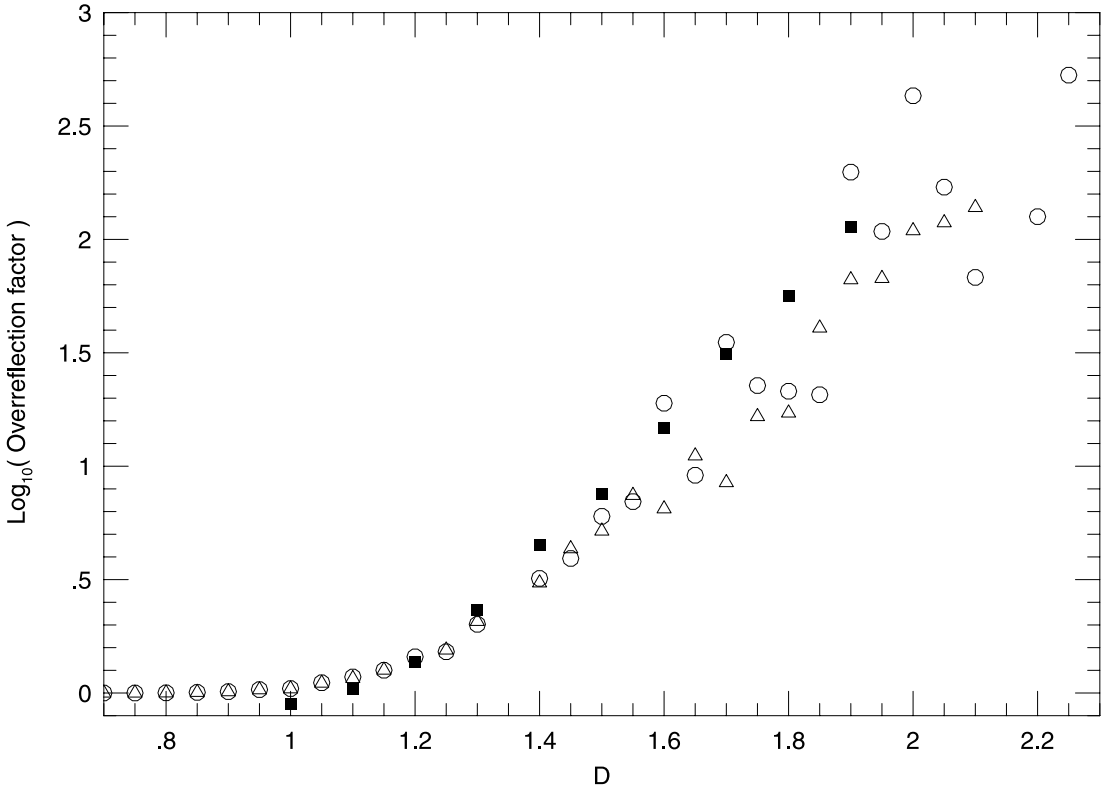

FIG. 5.—Overreflection factors as a function of the dimensionless rotational speed D. Circles and triangles correspond to separate estimates of the overreflection factor, as described in the text. Squares give the results from wave packet simulations of the type described in § 7. These simulations cannot be performed for low values of D without exciting the axisymmetric collapse instability.

5.6. *Physical Interpretation*

To obtain a physical interpretation of our results, we appeal to the concepts of the propagation of wave energy and angular momentum at the group velocity of spiral density waves in the modal approach (see, e.g., Bertin et al. 1989). These concepts are rigorously derivable only in the asymptotic approximation for tightly wound spirals (see, e.g., Toomre 1969; Shu 1970; Dewar 1972; Lynden-Bell & Kalnajs 1972; Goldreich & Tremaine 1979). The coupling of short leading and short trailing spiral waves near the corotation circle is too nonadiabatic a process to be *quantitatively* amenable to a WKBJ-like analysis since, in the swing mechanism, they involve fairly abrupt changes in the signs of wavenumbers that are large in magnitude. Nevertheless, the phenomenon of overreflection depends *qualitatively* only on the physically plausible notions of conservation of wave energy and angular momentum in the presence of tunneling, transmission, and reflection. The essential point is that spiral density waves of either leading or trailing type carry negative energy and angular momentum interior to corotation and positive energy and angular momentum exterior to corotation. When a leading spiral wave is incident on the corotation circle from either side of it, therefore, this leading spiral must overreflect into a trailing spiral wave if transmission of a trailing spiral occurs to the other side of corotation (see Fig. 12.4 of Shu 1992).

Corotation occurs where $\Omega(\varpi) = \omega/m$, i.e., at a radius $\varpi = mD\Theta^{1/2}a/\omega$, which is finite if $\omega/m > 0$, and which is infinite if $\omega/m = 0$. In a disk where Ω approaches infinity as a negative power of ϖ when $\varpi \to 0$ (e.g., a SID where $\Omega \propto \varpi^{-1}$), every density wave appears as if it effectively had a corotation circle at infinity, i.e., a pattern speed ω/m, which is essentially infinitesimal in comparison with the matter frequency Ω. Thus, for given D and m, all density waves that propagate close to the center of the disk look as though they have a corotation circle at infinity, i.e., a wave frequency ω that is essentially zero, even if it is not actually zero. Power-law disks thus have the interesting property of acting as notch filters to select a single, dimensionless wavenumber α for the propagation of spiral waves of given m and any frequency ω near the origin.

Consider now the overreflection of waves propagating inward in the outer cavity toward the corotation circle. (The "reciprocity theorem" of the next section demonstrates an equivalent result for waves launched from the inner cavity.) Tunneling must occur to accomplish a transmission of waves across the corotation circle if that circle is straddled by a Q-barrier. The larger the Q-barrier, the weaker the tunneling, and the weaker, then, is the corresponding overreflection. At some stage, the Q-barrier extends right to the origin, and no waves can tunnel to the inner cavity from the outer cavity. The accurate criterion for when this happens is obtained by setting $\alpha = 0$ in the exact equation (41) and not by setting the location of the inner edge of the Q-barrier at the origin, $\xi = L_- + \Delta = 0$, from the approximate equations (46) and (47). When the wavenumber α goes to zero, spiral density waves are evanescent, and tunneling and wave transmission to the interior ceases. Without tunneling and wave transmission, there is no overreflection. Without the wave amplification provided by overreflection, no pattern of standing waves (normal modes) can grow secularly in time. In other words, the system will have no linear instabilities if we replace continuous propagation in a medium that is infinite in $\ln \xi$, toward both positive and

negative values, with reflecting boundary conditions somewhere in the system (an easy circumstance to realize in realistic disks).

Section 4 showed that zero-frequency spiral waves of all m are evanescent if D^2 falls below the critical value, 0.5368, where α goes to zero for $|m| = 2$. We have argued above that the amplification of spiral disturbances depends solely on whether they can transmit (and therefore overreflect) waves across corotation, ultimately to reach the origin. The self-similar nature of a power-law disk implies that the overreflection factor for given disk properties (e.g., for given D above the critical value in a SID) is *independent* of the numerical value of the wave frequency ω of the input steady wavetrain. Thus, the calculations of § 5.5 did not need to specify a specific choice for ω as long as it is greater than zero. These results form the physical basis of the "surprising conclusion" reached by Lynden-Bell & Lemos (1993) that a continuum of disturbances become unstable simultaneously in their analysis of this problem.

It should be clear from our discussion that the self-similarity of disturbances of given m and arbitrary ω (once we renormalize the distance in terms of our ξ variable) does not hold for disks that are not of a power-law type. The more generally valid statement is that the stability of any disk against spiral disturbances the growth of which depends on swing amplification can be found by analyzing the transmission and tunneling properties of such disturbances across corotation. This statement is in agreement with the viewpoint expressed by Bertin et al. (1989).

5.7. *Dynamical Barred-Spiral Instability*

The form taken by modal eigenfunctions at the onset of spiral instability in modified or unmodified SIDs deserves note. As described by equation (54), near the origin, the perturbation surface density must be a superposition of

$$S(\varpi) \propto \varpi^{-3/2} \sin(\alpha \ln \varpi) \tag{65}$$

and its linearly independent counterpart

$$S(\varpi) \propto \varpi^{-3/2} \cos(\alpha \ln \varpi) . \tag{66}$$

For $\alpha \neq 0$, no linear combination of equations (65) or (66) can produce a perturbation surface density $S(\varpi)$ in the limit $\varpi \to 0$ that is equally or less divergent than the unperturbed surface density $\Sigma_0(\varpi) \propto \varpi^{-1}$. This breakdown of perturbation theory is cured in the analyses of Zang (1976) and Evans & Read (1998) by cutting out the central regions of the disk. However, for a SID at the margin of stability against spiral modes, $\alpha \to 0$ as $\varpi \to 0$ (see § 5.2). In this case, equation (65) represents the correct superposition of trailing and leading spirals, with a precisely defined phase relationship (90°) between the two upon "reflection" from the origin. In the double limit of marginally stable, nonaxisymmetric, "breathing modes," where $\alpha \to 0$ before $\varpi \to 0$, $S(\varpi)$ is then well-behaved, vanishing at the origin even without introducing any cut-outs (cf. Goodman & Evans 1999). To be sure, when D^2 exceeds the critical value 0.5368, such disks need to be modified appropriately in their central regions if growing $m = 2$ disturbances are to remain well behaved at small ϖ. In the numerical simulations of § 7 (see also Fig. 10), we give an explicit example of a growing mode that is barlike at small ϖ and spiral-like at large ϖ. For these reasons, we refer to this dynamically growing normal mode as a "barred-spiral instability." The word association with barred spiral galaxies is deliberate (see Shu 1992, chap. 12).

We propose therefore that $D^2 = 0.5368$ signals the onset of dynamical barred-spiral instabilities in appropriately modified but full SIDs and that the growths of these barred spirals are achieved by the same swing-amplification mechanism we have identified for ordinary-spiral instabilities. In this picture, barred spiral galaxies and ordinary spiral galaxies owe their morphological differences primarily to how density waves behave as they propagate toward the central bulges and nuclei of disk galaxies. We identify strongly barred spirals with inner feedback at the "breathing mode limit" of swing-amplified $m = 2$ spiral disturbances. This identification introduces a near-correspondence between the numerical criteria for secular and dynamical barlike distortions in compressible, differentially rotating, self-gravitating disks that has been noticed in many N-body simulations (cf. Ostriker & Peebles 1973). According to this analysis, the similarity between $D^2 = 0.5000$ and $D^2 = 0.5368$ for the two criteria of "barlike" instability in a full SID is largely a coincidence since the secular and dynamical modes possess completely different mechanisms of disturbance growth (see § 3.1): The secular bars or oval distortions arise from bifurcations of neighboring equilibria while the dynamical barred spirals arise from the limiting behavior of spiral disturbances at the centers of galaxies.

6. A RECIPROCITY THEOREM

In this section, we wish to show that for given D (or given Q), the overreflection factor for the inner cavity process, $c_+ \to c_-$, C_-, is the same as for the outer cavity process, $C_+ \to C_-$, c_-. For purposes of normalization, we adopt the convention that the inner wave amplitudes in the procedures, $c_+ : C_+^+$, C_-^+ and $c_- : C_+^-$, C_-^-, are unity, $|c_+^0| = |c_-^0| = 1$. The overreflection factor for the outer cavity process $C_+ \to C_-$, c_- is as described above but in a slightly augmented notation,

$$\mathscr{R}^- = \frac{|C_-^-|^2}{|C_+^-|^2} . \tag{67}$$

The inner cavity process $c_+ \to c_-$, C_- has inner cavity amplitudes, $|c_+| = |c_+^0| = 1$, $|c_-| = |Nc_-^0| = |N| = |C_+^+|/|C_+^-|$. Thus, the overreflection factor for the inner cavity process is

$$\mathscr{R}^+ = \frac{|c_-|^2}{|c_+|^2} = \frac{|C_+^+|^2}{|C_+^-|^2} . \tag{68}$$

Comparing the expressions for $\mathscr{R}^-$ and $\mathscr{R}^+$, we see that the only difference is the numerators, $|C_-^-|^2$ and $|C_+^+|^2$. But $|C_-^-|$ is the amplitude of the trailing-wave response in the outer cavity to a unit amplitude trailing wave in the inner cavity, whereas $|C_+^+|$ is the amplitude of the leading-wave response in the outer cavity to a unit amplitude leading wave in the inner cavity. The time-steady responses are invariant to time reversal. Time reversal changes the sign of the flow velocity without changing the perturbation density pattern that the flow produces (as required by the antispiral theorem of Lynden-Bell & Ostriker 1967); thus, time reversal changes leading spirals into trailing spirals of the same amplitudes, and the second situation becomes the same as the first. In particular, it requires $|C_+^+| = |C_-^-|$, i.e., $\mathscr{R}^+ = \mathscr{R}^-$.

The foregoing line of reasoning is quite general and does not rely on any property peculiar to a SID. Our conclusion follows mathematically by noting that the coefficients of equations (19) and (22) are real when ω is real. Thus, if S and Φ are solutions of equations (19) and (22), then so are their complex conjugates, S^* and Φ^*; but complex conjugation transforms leading spirals into trailing ones, and vice versa. This requires $|C_+^+| = |C_-^-|$.

The reciprocity relation applies to any disk for which we can identify "initial" and "final" states for the wave-wave interaction process. The name "reciprocity theorem" is inspired by quantum mechanics, in which a similar symmetry is proved for forward and backward scattering amplitudes (see, e.g., Baym 1973). The main difference between spiral-wave scattering and quantum scattering is the possibility for overreflection in the former; spiral waves allow the existence of states of both positive and negative densities of wave energy and angular momentum. Overreflection is forbidden in quantum mechanics if we rule out states with negative probability density.

The reciprocity theorem can be used to strike a compromise between the long-standing difference in interpretation of the cause of spiral structure in disk galaxies. In the modal hypothesis (Bertin et al. 1989), spiral structure is viewed as being primarily an internal instability phenomenon; in the swing-amplification hypothesis (Toomre 1981), spiral structure is thought to stem primarily from tidal interactions. The reciprocity theorem demonstrates that if a disk is sufficiently responsive as to swing-amplify outer-cavity disturbances created by a passing galaxy to a dramatic spiral pattern, then the target disk has access to the same swing amplification mechanisms for the growth of inner-cavity normal modes, and vice versa.

The validity of the spiral-wave reciprocity theorem depends on the absence of true resonances. In the presence of true resonances, it is necessary to adopt the Landau procedure of letting $\omega \equiv \omega_R + i\omega_I$ become real in the limit of vanishingly small growth rate $\omega_I \to 0-$. For example, such a procedure shows that corotation resonances (if the gradient of the vorticity divided by surface density $\equiv$ vortensity does not vanish at corotation) can act as emitters or absorbers of waves. Gas that moves slightly faster or slower than the wave on one side or the other can transfer energy and angular momentum to or from the wave (see Goldreich & Tremaine 1979).

In the WKBJ approximation, Goldreich & Tremaine (1979) show that, unlike a stellar dynamical disk (Shu 1970; Mark 1971), there are no such secular effects in gas dynamical disks associated with the Lindblad resonances. Thus, when ω is real and the gradient of the vortensity vanishes, our numerical treatment can simply avoid putting the Lindblad resonances at a grid point, a procedure that is essentially tantamount to taking only the principal value when integrating over the pole contributions, $(\xi - L_+)^{-1}$ or $(\xi - L_-)^{-1}$, that are the factors of the term $\mathscr{D}$ defined by equation (53). Provided the Goldreich and Tremaine analysis holds, the principal-value treatment captures all of the physics of our problem (because a simple pole gives canceling positive and negative contributions as one crosses it).

Feedback at the inner edge or origin is a way to convert inwardly propagating c_- into outwardly propagating c_+; feedback at an outer edge is a way to convert outwardly propagating C_- into inwardly propagating C_+. If only one type of reflection occurs, we have three-wave modal schemes (two waves on the side of the reflecting boundary; one wave on the side that acts as if it has radiation or absorption boundary conditions). In stellar dynamical disks, the latter can be justified on the basis of wave absorption at Lindblad resonances. In gas dynamical disks, they might be justified on the basis of dissipative effects that arise when waves reach nonlinear amplitudes. If both sides of corotation have reflecting boundary conditions, then we have to consider four-wave modal processes. In all cases, the "quantum condition" associated with reflecting the waves into the corotation region with the proper phase as to give coherent reinforcement of the original waves yields a discrete selection for the allowable eigenvalues of ω. The disk can support *disturbances* with a continuum of values of ω, and such *disturbances* will all be excited in some general initial-value process, but only the discrete *normal modes*, the traveling-wave components of which coherently reinforce each other, will grow in time.

7. TIME-DEPENDENT WAVE PACKETS

As the final study of this paper, we complement the steady wavetrain analysis by examining the time-dependent fate of various spiral wave packets in the disk. This approach, which gives a very pictorial depiction of the processes of wave propagation, tunneling and transmission, and overreflection, was introduced by Toomre (1969) in his investigation of the group velocity of spiral density waves.

Our simulations of wave packet propagation use the V2D.f hydrodynamics code (Laughlin 1994), which adopts a second-order van Leer advection scheme to integrate equations (5), (6), and (7), and a fast Fourier transform (FFT) algorithm (see Binney & Tremaine 1987) to solve for the gravitational potential. The simulations reported here have been done with 256 logarithmically spaced radial zones and 64 equally spaced azimuthal zones.

The computational grid covers 3 orders of magnitude in radial dynamic range, such that the ξ coordinate, defined for a wave frequency ω centered on those found in the packet, varies from an inner radius of $\xi = 0.02$ to an outer radius of $\xi = 20.0$. Because of softening introduced by the FFT gravity solver, and the finite radial range of the disk, it is necessary to introduce an extra radial force, $F_{\rm ext}$ to produce force balance in an unperturbed disk model specified by equation (9). In Figure 6, we plot the radial forces over the radial extent of the disk. At radii interior to $\xi \approx 0.1$, $F_{\rm ext}$ is quite large and will thus compromise the validity of the asymptotic solutions near the inner grid edge. Throughout most of the disk, $F_{\rm ext}$ is small in comparison to the other forces. The dashed vertical lines show the locations of the Lindblad resonances and the corotation radius for $m = 2$.

As a first example, we trace the evolution of an $m = 0$ wave packet in a $D = 1.6$ disk. The time evolution of the packet is shown in Figure 7. As expected, the group velocity causes the packet to split apart and propagate toward the edges of the disk. The waves reach the inner edge at approximately $t = 6.0$, and are fully reflected by $t = 9.0$, explicitly demonstrating the point made by Goodman & Evans (1999).

In the following simulations, we employ a sponge boundary condition designed to minimize reflection of wave energy at the inner radius of the disk. For grid cells interior to $\xi = 0.1$, we gradually enforce an admixture of the equi-

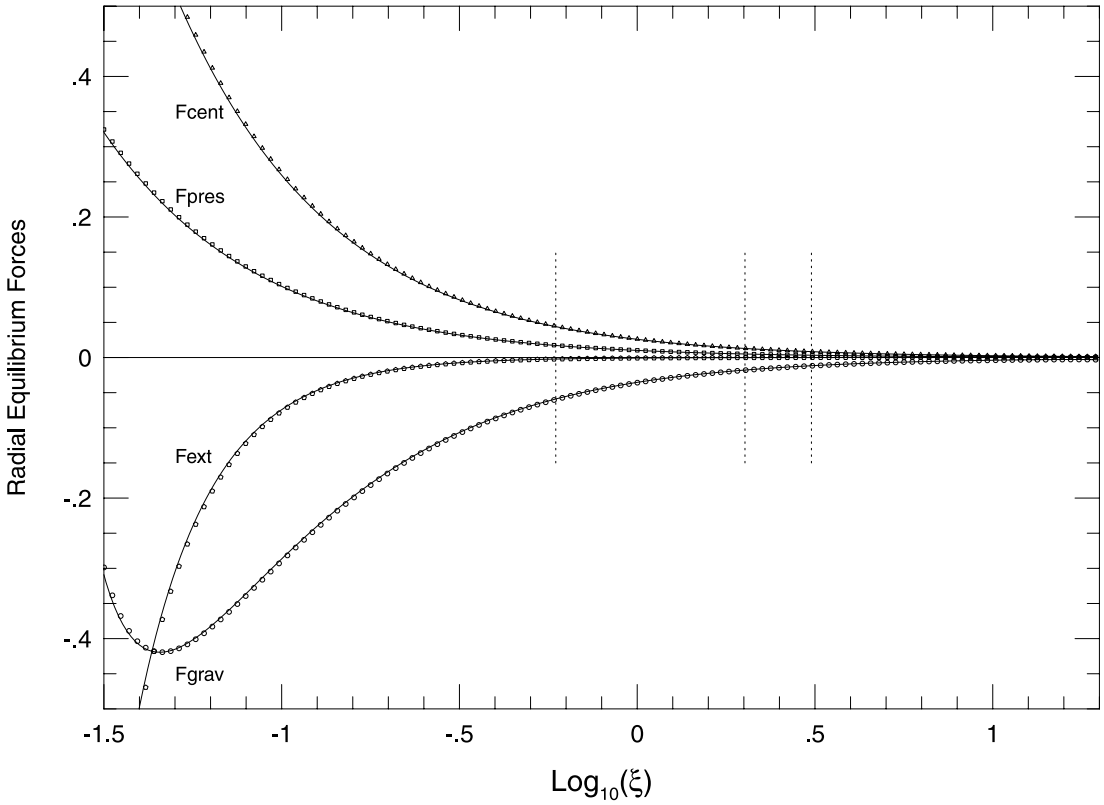

FIG. 6.—Distribution of the centrifugal, pressure, self-gravitational, and "external" forces for the equilibrium hydrodynamical model when $D = 1.6$. The artificial external term is needed to restore radial force balance because of the disk truncation and gravity softening introduced in the numerical treatment. Small circles show forces obtained with 128 radial grid points. Solid lines connect the model obtained with 256 radial grid points. The dotted vertical lines show the ILR, CR, and OLR for subsequent perturbation simulations.

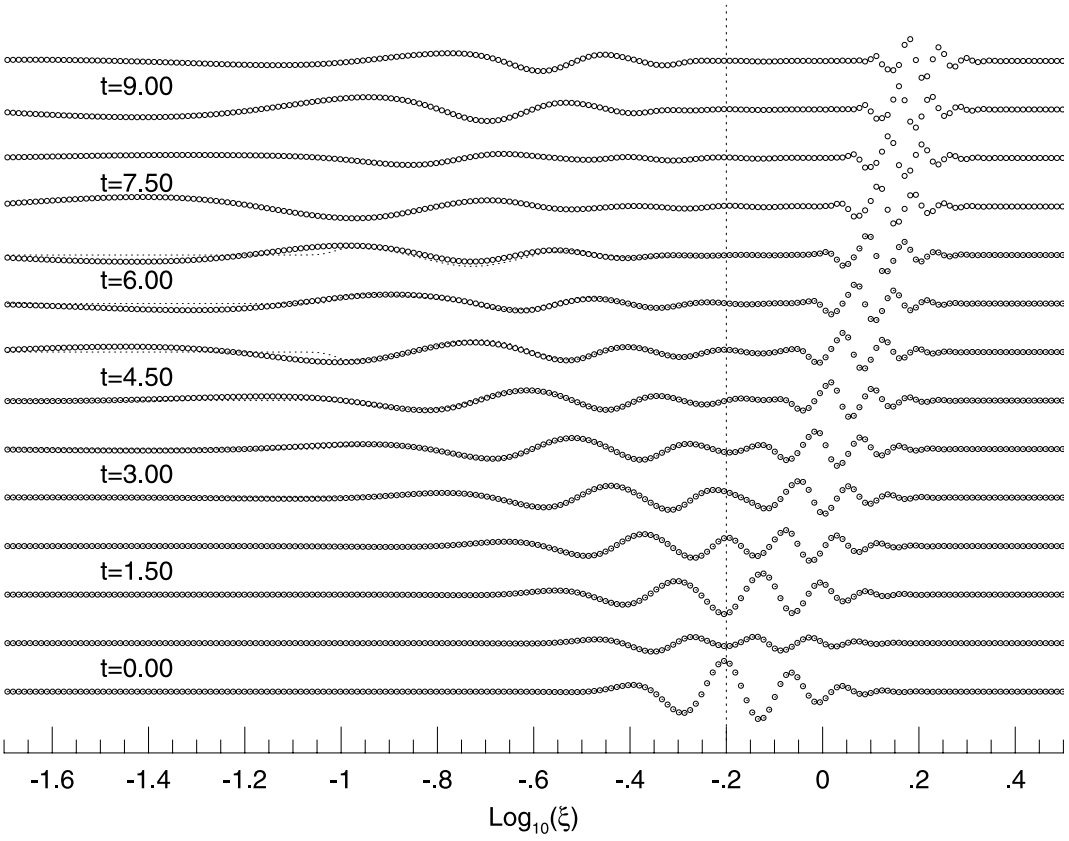

FIG. 7.—Propagation of a stable $m = 0$ disturbance in a $D = 1.6$ disk. Departures from the equilibrium surface density profile are plotted at equally spaced intervals in time. The dashed line marks the peak of the initial disturbance.

librium solution given by equation (9). At the inner edge of the disk, the fluid retains its initial equilibrium configuration.

Figure 8 shows the evolution of a wave packet of dimensionless wavelength 0.48 in a disk with $D = 3.0$. As shown by the graph of ω^2, which is superimposed in solid lines on Figure 8, the dispersion relation (eq. [39]) predicts that disturbances of wavelength 0.48 should be unstable in the region between $\log_{10}(\xi) = -0.75$ and $\log_{10}(\xi) = -0.25$. The time evolution of the packet in Figure 8 shows that this is indeed the case.

Wave-packet simulations can also be used to study the overreflection of propagating spirals. For example, to obtain a graphical depiction of the $C_+ \rightarrow C_-, c_-$ process, we apply a transient potential perturbation (shutting it off after $t = t_0 = 10./D$),

$$\mathscr{V}_{1_{\rm ext}}(\xi, \varphi, t) = 0.025\ \mathscr{V}_0(\xi) e^{-(\xi-\xi_0)^2/h^2} \times \sin(\pi t/t_0) e^{i(D\xi+\omega t-m\varphi)}, \quad (69)$$

to the disk. This perturbation produces a leading spiral wave packet that has a radial wavelength appropriate to the asymptotic outer cavity for the pattern speed ω/m. The fate of such a packet (with $m = 2$, $h = 2.7$, and $\xi_0 = 7.7$) is shown in Figure 9, the five panels of which cover two full pattern revolutions for a $D = 1.6$ disk. The evolutionary trend is from bottom to top. The panels on the left show the real and imaginary components of the two-armed disturbance in the disk (scaled by $\xi^{3/2}$). The panels on the right show the spiral pattern within the radius $\xi = 10.0$. The strength of the spiral at a particular radius is indicated by the size of the dots.

The earliest figure in the time sequence shows the disturbance in the disk soon after the potential perturbation has ended. The leading spiral propagates inward in the prograde sense. In addition, the initial perturbation also generates an outward propagating, retrograde leading spiral. In subsequent frames, one can trace the development of the inwardly propagating spiral as it undergoes overreflection off the corotation radius.

The steady wavetrain analysis indicates that the overreflection factor for this $D = 1.6$ disk should be roughly 10. To measure the overreflection factor for the wave packet in the simulation, we have charted the $m = 2$ Fourier amplitude at the radius $\xi = 3.16$ ($\log_{10}\xi = 0.5$) in the disk. Figure 10 plots the $m = 2$ Fourier amplitude at $\xi = 3.16$ versus "radial wavenumber" $\alpha/\xi \equiv m\, d\varphi/d\xi$ measured along a locus of constant spiral phase, as the simulation progresses. At $t = 44.75$, the amplitude of the inbound leading wave reaches a maximum of 0.05. Soon thereafter, the leading spiral swings into a trailing spiral. At $t = 96$, the peak of the outgoing wave packet crosses $\xi = 3.16$ with amplitude 0.17. The overreflection factor is thus $(0.17/0.05)^2 = 11.6$, which is similar to the value obtained with the matrix method.

The dispersive properties of wave packets complicate the measurement of the overreflection factors of the individual wavetrains that superimpose to make a wave packet. In a SID, all prograde spiral wavetrains of a given m have the

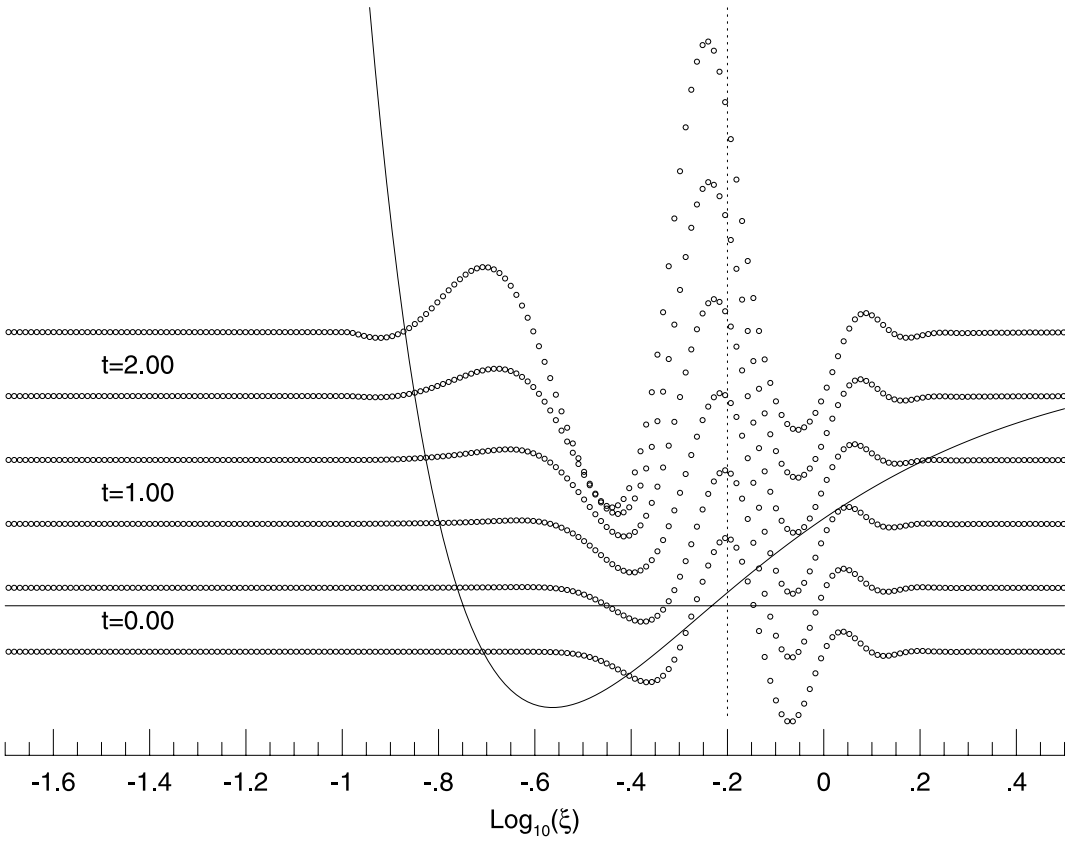

FIG. 8.—Growth of an unstable ringlike perturbation in a $D = 3.0$ ($Q = 0.8485$) disk. The solid curve shows the square of the wave frequency, ω^2, evaluated from the WKBJ dispersion relation in equation (39) for an axisymmetric wave of the properties shown at $t = 0.00$. Asymptotic theory predicts that the disturbance should grow exponentially with time where ω^2 falls below zero (*solid horizontal line*) and propagate as an oscillatory disturbance where ω^2 rises above zero. Both expectations are borne out by the sequence of curves delineated by small open circles showing the simulation disturbance at equally spaced intervals in time.

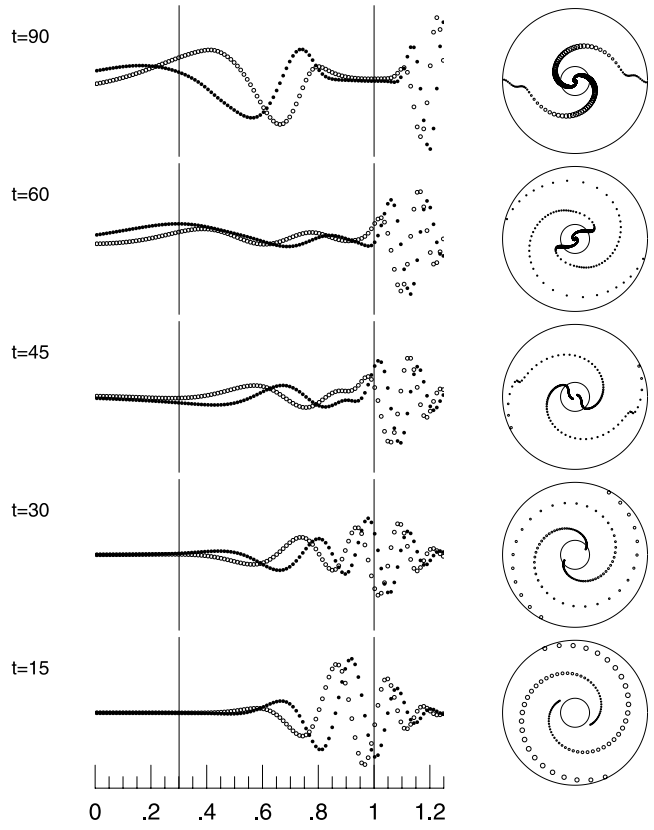

FIG. 9.—Moderate overreflection of an initially leading-spiral wave packet in a $D = 1.6$ disk.

same overreflection factors, so a power-spectrum analysis (which we did not carry out) should be insensitive to this complication. A simple amplitude analysis of the type presented could suffer from chance constructive or destructive interference.

Another complication also enters in our time-dependent simulations. At later times an $m = 2$ barred-spiral mode is growing, and the measured radial wavenumber formally approaches zero. Added to the experience of many other numerical simulations in the literature, this result suggests that the introduction of length scales (here, by the numerical procedure) may allow dynamically unstable barred-spiral modes to develop in appropriately modified SIDs that rotate rapidly enough, even though only the secularly unstable aligned disturbances belonging to a different family are revealed by a linearized perturbation analysis of the unmodified SID. In particular, despite the fact that the present simulations adopt "sponge boundary conditions" to soak up inwardly propagating trailing-spiral waves, the absorption is imperfect since equation (65) implies that vanishing perturbations as $\varpi \to 0$ constitute the correct superposition of leading and trailing spirals of zero wavenumber α. Thus, some reflection into leading spirals occurs near the origin, establishing the resonant cavity needed for a rapidly growing mode that is barlike in the central regions and spiral-like in the outer regions (see circular inset of Fig. 10). (Unfortunately, we are unable to study by the present simulation techniques the behavior of the dynamical barred-spiral mode in disks closer to the putative margin of stability, $D^2 = 0.5368$, because such full SIDs are violently

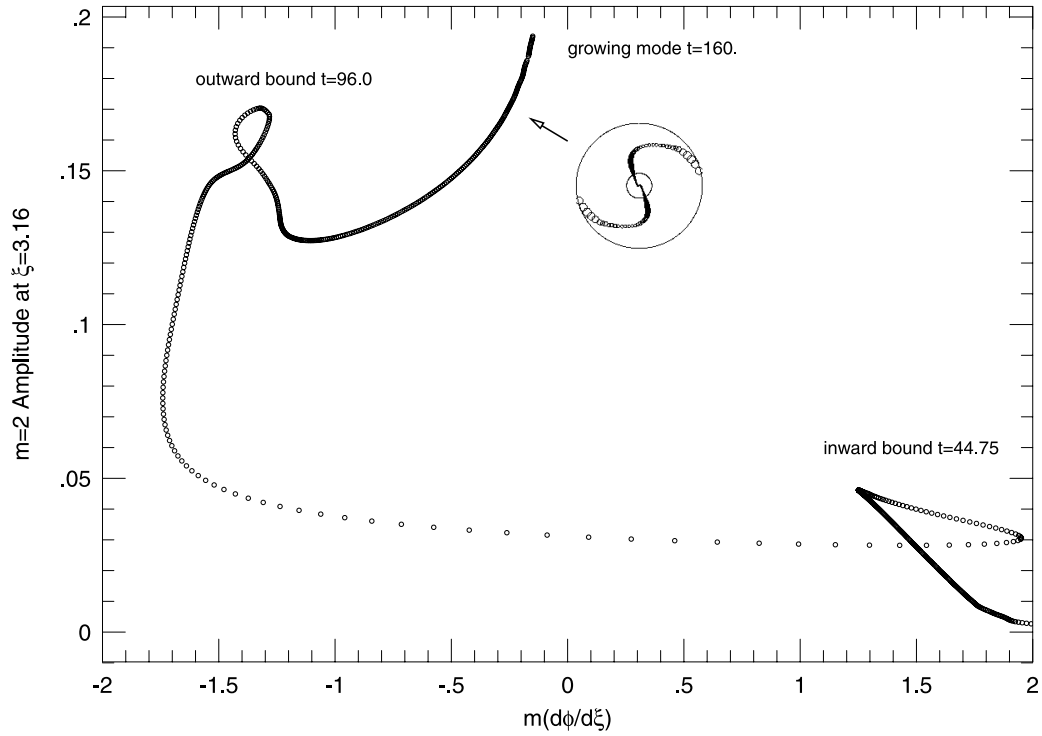

FIG. 10.—Wave amplitude vs. radial wavenumber $\alpha/\xi \equiv m\,d\varphi/d\xi$ as measured by a loci of constant spiral phase during a $D = 1.6$ spiral wave-packet simulation with $m = 2$. Because the disk is also unstable to an aligned $m = 2$ oval distortion, at late times, $t \sim 160$, the simulation is dominated not by the overreflection of the input spiral waves but by a growing barlike structure. The structure of the dynamical bar instability, together with the spiral arms that it forces, is shown in the circular inset.

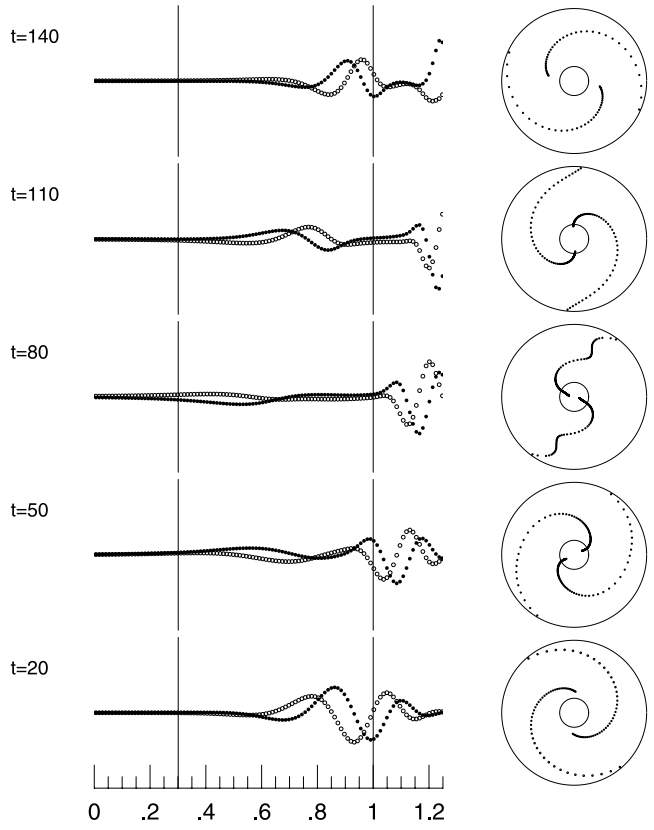

FIG. 11.—Slight overreflection of an initially leading, two-armed, spiral-wave packet in a $D = 1.0$ disk.

unstable to axisymmetric collapse; see also the caption to Fig. 5.)

Figure 11 illustrates the overreflection process for a $D = 1.0$ disk. In this case, the overreflection factor (measured at $\xi = 5.0$) is 0.89, as compared with a value of 1.05 predicted by the steady wavetrain analysis. A summary of the overreflection factors found by such wave packet simulations is given in Figure 5 (*squares*).

8. DISCUSSION

The simplicity inherent in self-similar disks of power-law form has allowed us in this paper to obtain a unifying view of much of the work done on spiral density waves over the last 35 years. From this perspective, we see that, contrary to the general perception, there is *no* phenomenon revealed by exact global analysis that cannot be understood, at least *qualitatively*, by the judicious application of WKBJ theory. To be sure, the *quantitative* calculation of overreflection factors (or growth rates in a normal-mode analysis) is difficult to achieve using standard asymptotic methods when swing amplification underlies the wave growth because the swing interaction across corotation of short leading and short trailing waves does not involve the adiabatic evolution of one waveform to another. Nevertheless, the WKBJ concepts of group propagation, localization of wave energy and angular momentum of opposite signs, and barrier tunneling are all extremely helpful—perhaps indispensable—to a *physical* understanding of the important wave dynamics.

We hope that our proof of the reciprocity theorem in § 6 will also have a salutary liberating effect on the field. The long-standing controversy over whether spiral structure in disk galaxies is created by internal modal instabilities or by external forcing will perhaps now be approached with greater equanimity. The reciprocity theorem states that if one outlook is possible, the other necessarily also has validity. Galaxies capable of swing amplifying external disturbances into dramatic spiral structures necessarily have access to the same growth mechanisms for internal spiral modes, and vice versa. Which process is dominant in any given case will depend on the important details of the problem: on the availability of massive companions orbiting appropriately with respect to the target galaxy and the existence of internal reflecting boundaries that can establish a resonant cavity with an amplifier at the corotation circle. In all likelihood, as C. C. Lin has emphasized for many years, there is a coexistence of the two principal mechanisms that produce the beautiful structures that astronomers observe in the universe of spiral galaxies.

Similarly, we hope that we have made plausible the speculation that barred-spiral structure is simply a limiting example of ordinary-spiral structure and that both types of normal modes use the same basic mechanism for disturbance growth: swing amplification by overreflection across the corotation circle. Thus, supporters of the position that spiral arms in galaxies can be driven by central bars have no logical reason to object any longer to the opinion that ordinary spirals can also owe their grand designs to modal instability.

We have also shown that the aligned bifurcations discovered in scale-free disks by Syer & Tremaine (1996), associated with $m = 1$ eccentric displacements, $m = 2$ oval distortions, etc., are of a fundamentally different character than the spiral disturbances (barred and ordinary). First, the zero-frequency aligned disturbances at finite amplitude (discussed in Paper II, forthcoming) represent adjacent equilibria of small energy separations (except for $m = 1$) and with a constant or nearly constant ratio of circulation to enclosed mass, which are likely to be accessed by starting conditions that are axisymmetric only in the presence of unconsidered dissipative mechanisms (most promisingly in the star formation context, ambipolar diffusion or turbulent dissipation). Since dissipative processes are believed to be responsible for molecular cloud-core formation, there exists motivation to consider the possibility that the pivotal starting states for dynamical collapse calculations might be nonaxisymmetric rather than axisymmetric. Numerically, a dimensionless square of the rotation rate $D^2 < 0.5386$ suffices to suppress overreflection of spiral density waves of all m and ω in a full SID. But $m = 2$ oval distortions can be present until D^2 becomes less than 0.5000, and $m = 1$ eccentric displacements can occur at all values of $D^2 \geq 0$. Molecular cloud cores commonly rotate slower than $D \sim 0.7$ times their magnetosonic speed, so they are dynamically stable against spiral disturbances of any $|m| \geq 1$ and secularly stable against aligned perturbations with $|m| \geq 2$, but they are not secularly stable against aligned $|m| = 1$ eccentric displacements. Moreover, magnetized gaseous configurations differ from unmagnetized ones in that the former can be fairly flattened even if they are slowly rotating. If such flattened objects resemble singular isothermal disks, like the pivotal states to which theories of ambipolar diffusion show them to evolve (Nakano 1979; Lizano & Shu 1989; Basu & Mouschovias 1994; Li & Shu 1996), then the centers of such cores might prefer to become eccentrically displaced with respect to the ambient material. Perhaps this

phenomenon explains why young stellar objects, such as protostars and T Tauri stars, are rarely, if ever, found at the centers of the surrounding isodensity contours of molecular matter (P. Myers, 1998, private communication).

Bifurcations of axisymmetric, rotating, self-gravitating configurations into nonaxisymmetric forms appeared first in the classic studies of incompressible ellipsoidal figures of equilibrium (see Jeans 1928; Chandrasekhar 1969; Tassoul 1978). Many studies have shown that similar bifurcations do not occur in uniformly rotating gaseous configurations of sufficient central concentration (e.g., polytropes; see James 1964), although they can arise in compressible equilibria that are sufficiently differentially rotating (e.g., Ostriker & Mark 1968; Bodenheimer & Ostriker 1973). However, until the pioneering work of Syer & Tremaine (1996), there have been few actual constructions of self-consistent states of nonaxisymmetric, self-gravitating, compressible, differentially rotating, equilibria. In Paper II (forthcoming), we amend this deficiency in the case of (magnetized) singular isothermal disks by giving a simple systematic technique for finding such equilibrium states for any realizable amplitude of inviscid distortion and for any m-fold azimuthal symmetry. These two-dimensional results represent a first attempt to model pivotal states of molecular cloud cores that are not postulated to be perfectly axially symmetric.

Even if molecular cloud cores start out slowly rotating and axisymmetric when they reach their pivotal states, it is possible to show that they are likely to lose that axial symmetry in the subsequent dynamical collapse. If a slowly rotating SID collapses isothermally and axisymmetrically, without any redistribution of specific angular momentum (equivalent to circulation in an axisymmetric configuration) or magnetic flux, then we can easily demonstrate that it will collapse into a more rapidly rotating SID with the same values of a, ϵ, and Θ as enter in equation (9), but with a final rotation parameter D_f that is related to the initial (pivotal) one D_i by[1]

$$D_f D_i = 1 \ . \tag{70}$$

Thus, an initial cloud core rotating slowly enough, $D_i < 1/\sqrt{2}$, to be secularly stable with respect to all $|m| \geq 2$ aligned disturbances and dynamically stable to all $|m| \geq 1$ spiral perturbations, will be dynamically unstable to axisymmetric collapse.[2] In the absence of any redistribution of specific angular momentum or magnetic flux (e.g., in the absence of viscous mechanisms that can act on dynamical timescales), this collapse will result in a starless protostellar disk that is rotating rapidly enough, $D_f > \sqrt{2}$, to resist further radial contraction but will be dynamically unstable in turn to infinitely many spiral disturbances (see Table 1). It is this almost inevitable departure from axial symmetry at some stage of the star formation process that bodes well theoretically for the prospects of binary- and multiple-star outcomes. To be sure, even if nonaxisymmetric, self-gravitating configurations are unavoidable, it remains to be seen under what conditions such instabilities will result in gravitational fragmentation into multiple stellar bodies and under what others will result an inward transport of mass and an outward transport of angular momentum that leads to the formation of a single star plus a planetary system.

At the time of the original submission of this article, the authors were unaware of the important paper by Goodman & Evans (1999). Because of the confusion in the literature concerning the existence and nature of normal modes in unmodified power-law disks, the referee encouraged us to try to clarify the conflicting viewpoints. We have attempted to do this in the present revision, and we thank the referee for many penetrating comments that made our job easier. This work was funded under the auspices of a special NASA astrophysics theory program which supports a joint Center for Star Formation Studies at NASA-Ames Research Center, University of California at Berkeley and at Santa Cruz. We dedicate this paper to the memory of James Mark, who pioneered the concept of the spatial overreflection of steady wavetrains across corotation.

[1] From equation (9) we compute that the annulus of radius ϖ and thickness $d\varpi$ contains mass $dM_0 = 2\pi\varpi\Sigma_0(\varpi)d\varpi = (\Theta a^2/\epsilon G)(1 + D^2)d\varpi$ and specific angular momentum $dj_0 = d(\varpi^2\Omega) = \Theta^{1/2}aD\ d\varpi$. If the distribution of mass with specific angular momentum is preserved in the collapse, the quantity $dM_0/dj_0 = (\Theta^{1/2}a/\epsilon G)(1 + D^2)/D$ remains invariant. In the presence of field freezing, ϵ and Θ are conserved, while a remains constant during isothermal collapse. Thus, we require $(1 + D_i^2)/D_i = (1 + D_f^2)/D_f$, which yields either the trivial solution $D_f = D_i$ or equation (70).

[2] Equation (70) seemingly implies that a SID can collapse axisymmetrically provided its initial D were less than unity, whereas the analysis of § 4.1 found instability only if D_i were less than $\sqrt{0.9320} = 0.9654$. The slight disagreement arises because the collapse under infinitesimal perturbations does not occur at once for the entire disk but only for a portion of it at any time t, from inside-out (see Li & Shu 1997). Only asymptotically in time $t \to \infty$, for a disk with $D_i < 0.9654$, would the entire SID be recreated with D_f satisfying equation (70).

REFERENCES

Adams, F. C., Ruden, S. P., & Shu, F. H. 1989, ApJ, 347, 959
Balbus, S. A., & Hawley, J. F. 1991, ApJ, 376, 214
Bardeen, J. M., Friedman, J. L., Schutz, B. F., & Sorkin, R. 1977, ApJ, 217, L49
Basu, S., & Mouschovias, T. 1994, ApJ, 432, 720
Baureis, P., Ebert, R., & Schmitz, F. 1989, A&A, 225, 405
Baym, G. 1973, Lectures on Quantum Mechanics (Menlo Park, CA: Benjamin Cummings)
Bertin, G., & Lin, C. C. 1996, Spiral Structure in Galaxies (Cambridge: MIT Press)
Bertin, G., Lin, C. C., Lowe, S. A., & Thurstans, R. P. 1989, ApJ, 338, 78
Binney, J., & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
Bodenheimer, P., & Ostriker, J. P. 1973, ApJ, 180, 159
Chandrasekhar, S. 1969, Ellipsoidal Figures of Equilibrium (New Haven: Yale Univ. Press)
Crane, P., et al. 1993, AJ, 106, 1371
Dewar, R. L. 1972, ApJ, 174, 301
Durisen, R. H., Gingold, R. A., Tohline, J. E., & Boss, A. P. 1986, ApJ, 305, 281
Evans, N. W., & Read, J. C. A. 1998, MNRAS, 300, 106
Goldreich, P., & Lynden-Bell, D. 1965, MNRAS, 130, 125
Goldreich, P., & Tremaine, S. 1978, Icarus, 34, 240
———. 1979, ApJ, 233, 857
Goodman, J., & Evans, N. W. 1999, MNRAS, 309, 599
Hayashi, C., Narita, S., & Miyama, S. M. 1982, Prog. Theor. Phys., 68, 1949
Hohl, F. 1971, ApJ, 168, 343
Hunter, C. 1977, ApJ, 213, 497
Huntley, J. M., Sanders, R. H., & Roberts, W. W. 1978, ApJ, 221, 521
James, R. A. 1964, ApJ, 140, 552
Jeans, J. H. 1928, Astronomy and Cosmogony (Cambridge: Cambridge Univ. Press)
Julian, W. H., & Toomre, A. 1966, ApJ, 146, 810
Kalnajs, A. J. 1971, ApJ, 166, 275
———. 1972, ApJ, 175, 63

Kalnajs, A. J. 1973, Proc. Astron. Soc. Australia, 2, 174
Kawabe, R., Ishiguro, M., Omodaka, T., Kitamura, Y., & Miyama, S. M. 1993, ApJ, 404, L63
Königl, A., & Pudritz, R. E. 2000, in Protostars and Planets IV, ed. V. Mannings, A. Boss, & S. Russell (Tucson, AZ: Univ. Arizona Press), in press
Laughlin, G. 1994, Ph.D. thesis, Univ. California, Santa Cruz
Laughlin, G., & Bodenheimer, P. 1994, ApJ, 436, 335
Laughlin, G., Korchagin, V. I., & Adams, F. C. 1998, ApJ, 504, 945
Lay, O. P., Carlstrom, J. E., Hills, R. E., & Phillips, T. G. 1994, ApJ, 434, L75
Lee, E., & Goodman, J. 1999, MNRAS, 308, 984
Lemos, J. P. S., Kalnajs, A. J., & Lynden-Bell, D. 1991, ApJ, 375, 484
Li, Z.-Y., & Shu, F. H. 1997, ApJ, 475, 237
———. 1996, ApJ, 472, 211
Liapunov, A. M. 1905, Memoirs St. Petersburg Acad., XVII, 3
Lin, C. C., & Lau, Y. Y. 1979, Studies Appl. Math., 60, 97
Lin, C. C., & Shu, F. H. 1966, Proc. Natl. Acad. Sci., 55, 229
———. 1968, in Proc. 1968 Summer Institute of Theoretical Physics, Brandeis University, ed. J. Goldstein (New York: Gordon & Breach), 235
Lizano, S., & Shu, F. H. 1989, ApJ, 342, 834L
Lynden-Bell, D., & Kalnajs, A. J. 1972, MNRAS, 157, 730
Lynden-Bell, D., & Lemos, J. P. S. 1993, astro-ph/9907093
Lynden-Bell D., & Ostriker, J. P. 1967, MNRAS, 136, 293
Mark, J. W.-K. 1971, ApJ, 169, 455
———. 1976, ApJ, 205, 363
Mestel, L. 1963, MNRAS, 126, 553
Miller, R. H. 1971, Ap&SS, 14, 73
Nakano, T. 1979, PASJ, 31, 697
Ostriker, J. P., & Mark, J. W.-K. 1968, ApJ, 151, 1075
Ostriker, J. P., & Peebles, P. J. E. 1973, ApJ, 186, 467
Pannatoni, R. F. 1979, Ph.D. thesis, MIT Univ.
Poincaré, H. 1885, Acta Math. VII, 259
Papaloizou, J. C., & Lin, D. N. C. 1989, ApJ, 344, 645
Schmitz, F. 1990, A&A, 230, 509
———. 1988, A&A, 200, 120
———. 1986, A&A, 169, 171
Shu, F. H. 1970, ApJ, 160, 99
———. 1977, ApJ, 214, 488
———. 1984, in Planetary Rings, ed. A. Brahic & R. Greenberg (Tucson: Univ. Arizona Press), 513
———. 1992, The Physics of Astrophysics, Vol. II: Gas Dynamics (Mill Valley, CA: University Science Books)
Shu, F. H., Allen, A., Shang, H., Ostriker, E. C., & Li, Z. Y. 1999, in The Origins of Stars and Planetary Systems, ed. C. J. Lada & N. D. Kylafis (Dordrecht: Kluwer), 193
Shu, F. H., & Li, Z.-Y. 1997, ApJ, 475, 251
Shu, F. H., Tremaine, S., Adams, F. C., & Ruden, S. P. 1990, ApJ, 358, 495
Skrutskie, M. F., Snell, R. L., Strom, K. M., Strom, S. E., Edwards, S., Fukui, Y., Mizuno, A., Hayashi, M., & Ohashi, N. 1993, ApJ, 409, 422
Syer, D., & Tremaine, S. 1996, MNRAS, 281, 925
Tassoul, J. 1978, Theory of Rotating Stars (Princeton: Princeton Univ. Press)
Toomre, A. 1964, ApJ, 139, 1217
———. 1969, ApJ, 158, 899
———. 1977, ARA&A, 15, 437
———. 1981, in Structure and Dynamics of Normal Galaxies, ed. S. M. Fall & D. Lynden-Bell (Cambridge: Cambridge Univ. Press), 111
———. 1982, ApJ, 259, 535
Zang, T. 1976, Ph.D. thesis, MIT Univ.